UniOfGalway@IberLEF 2024:Hope Speech Recognition in Spanish: A Comparative Analysis of Transformer-Based Models

Arunraj Subburaj[1,*,†], Amirthagadeshwaran Kathiresan[1,†], Rahul Ponnusamy[2], Paul Butilaar[2] and Bharathi Raja Chakravarthi[3]

[1]*University of Galway, Ireland*

[2]*Data Science Institute, University of Galway, Ireland*

[3]*School of Computer Science, University of Galway, Ireland*

Abstract

Hope speech detection in social media is crucial for fostering positive engagement and resilience within marginalized communities. By utilizing machine learning and NLP techniques, researchers have focused on identifying hope speech in online content, achieving F1-score between 76% and 84%. This detection of hope speech contributes to understanding how social media platforms can either create a hopeful or hostile environment, influencing community sentiments and individual identities. Recognizing and promoting hope speech is essential for enhancing interaction quality and support among users, especially within Equality, Diversity, and Inclusion initiatives, emphasizing the significance of hope in digital communication for building inclusive and supportive online ecosystems.

1. Introduction

Hope is a significant psychological construct that influences human emotions, behavior, and overall well-being. Individuals with high levels of hope tend to exhibit resilience in the face of challenges, viewing them as opportunities for growth, which can lead to positive outcomes such as academic success and reduced levels of depression [1]. Despite the profound impact of hope on individuals, its exploration within the realm of Natural Language Processing (NLP) has been limited [2]. Recent efforts have been made to promote research on hope through shared tasks, particularly focusing on hope speech detection in various domains, including contexts related to Equality, Diversity, and Inclusion (EDI) [3].

Social media platforms have become integral in shaping human behavior and decision-making processes, with a growing number of individuals engaging with these platforms [4]. The increased usage of social media has provided opportunities for informed decision-making based on social media sentiment, particularly influencing marginalized groups such as women in STEM fields, the LGBTQ+ community, racial minorities, and individuals with disabilities. While social media serves as a vital source of support and affirmation for many minority groups, it also presents risks, especially for young internet users [4].

In the computational domain, various methods have been developed to identify hope speech, distinguishing it from neutral or non-hopeful content. Techniques such as deep learning, transformer models,

and linguistic features have been employed for hope-speech detection, emphasizing the importance of online interactions in shaping individuals' identities and perceptions of society [5]. Social media platforms often exhibit two prevalent tones: hope and hate, with hope speech characterized by a positive outlook anticipating favorable outcomes. Detecting hope speech in social media is crucial as it provides comfort and empathy for marginalized groups seeking relatable content [6].

Hope plays a fundamental role in human psychology, impacting emotions, behavior, and well-being. Leveraging NLP techniques to analyze social media data for hope speech detection can offer valuable insights into human behavior, decision-making, and the promotion of positive and inclusive online environments, particularly for vulnerable communities like the LGBTQ+ community. Further research in this area is essential to harness the potential of hope in fostering well-being and resilience among diverse populations.

This paper presents a novel deep-learning method for detecting hope speech and identifies effective psycho-linguistic and linguistic features for the Spanish language. Section 2 details the Literature Review, Section 3 describes the task and dataset statistics, while Section 4 outlines the methodology employed. Sections 5 and 6 discuss the results and conclusions derived from the study.

2. Literature Review

Hope speech detection is a critical area of research, particularly in the realm of social media and its impact on minority communities. Several studies have delved into this area, emphasizing the importance of leveraging advanced technologies like deep learning and transformer-based models for effective hope speech detection. Palakodety et al. [7] provide a foundational approach to hope speech detection, their model's limited linguistic diversity raises questions about its applicability across different cultural contexts. This study extends their framework by incorporating a wider range of linguistic and cultural nuances, particularly focusing on the Spanish-speaking LGBTQI community, thus addressing the noted limitations. Additionally, Balouchzahi et al. [3] and Deepanshu et al. [6] demonstrated a robust methodology utilizing BERT embeddings for hope speech detection. However, their study did not explore the impact of informal language and slang frequently used in social media, which could affect the generalizability of their model. Our study addresses this gap by incorporating an enhanced preprocessing step that adapts to informal language variations, thereby improving the model's performance across diverse social media texts.. These studies underscore the importance of utilizing advanced technologies for effective hope speech detection [8].

Hope recognition in Spanish has gained significant attention in recent shared tasks such as LT-EDI-EACL and Iber-LEF. Various transformer-based models have been developed to detect hope speech in social media content. Transformer-based models, such as mBERT and XLM, have demonstrated remarkable performance in cross-lingual understanding tasks [9]. These models utilize large Transformer models pre-trained on multiple languages, showcasing their effectiveness in handling multilingual data. Additionally, the combination of XLM-RoBERTa language models has proven successful in detecting hope speech in social media [6]. Teams like giniUs and LPS have contributed to this field by presenting transformer-based models for hope speech detection, emphasizing the importance of identifying hopeful comments for equality, diversity, and inclusion [10]. The shared tasks focus on automatically identifying hopeful comments to promote positive sentiments and well-being in online interactions by LT-EDI-EACL 2022 [11]. The significance of hope speech detection for equality, diversity, and inclusion is underscored in various shared tasks and workshops [12, 13, 14]. These initiatives aim to identify hope speech in comments to foster positivity and support among individuals. Furthermore, the utilization of Transformer models in speech recognition systems has been widely recognized for their capacity to parallelize tasks and integrate internal attention mechanisms [15, 16].

Furthermore, the research by Chakravarthi et al.[17] showcases the success of models in identifying and eliminating negativity, demonstrating the potential of automated systems in fostering positive discourse. Moreover, Sidorov et al. [18] and García-Baena et al. [19] explore the effectiveness of transformer models in detecting hope speech within the LGBT community. While their findings are

promising, they fall short in addressing the dynamic nature of language evolution on social media platforms.It also emphasize the effectiveness of transformer models and contextual-aware resources in detecting hope speech, especially in multilingual contexts like Spanish and English.

Theoretical frameworks supporting hope speech detection often involve sentiment analysis and hate speech detection. Studies by Laaksonen et al. [20] and Zhu [12] discuss the datafication of hate speech and the importance of hate speech detection in society. These frameworks provide a foundation for understanding the context in which hope speech detection operates, highlighting the necessity of distinguishing between positive and negative speech content.

The research landscape on hope speech detection is diverse, encompassing deep-learning models and multilingual approaches, all aimed at promoting positivity and inclusivity in online dialogues. By leveraging advanced technologies and theoretical frameworks from related fields such as sentiment analysis and hate speech detection, researchers are making significant progress in identifying and promoting hope speech on social media platforms.

3. Task and Dataset Description

Hope speech detection within the LGBTQI community and across various domains is a crucial task with practical applications that necessitate models to generalize effectively. Numerous studies have investigated related areas, offering valuable insights for this task. The Language Technology for Equality, Diversity, and Inclusion (IberLEF-2024) dataset was utilized in the IberLEF 2024 shared task [21, 22, 19, 8, 18] on Hope Speech Detection for Equality, Diversity, and Inclusion. The IberLEF-2024[1] dataset plays a significant role in advancing language technology applications that support these important societal goals.

The datasets used in this study were sourced from the Language Technology for Equality, Diversity, and Inclusion (IberLEF-2024) dataset, made available by IberLEF 2024[2] for the shared task on Hope Speech Detection [21, 22, 19, 8, 18]. This dataset comprises 1,400 labeled tweets and an additional 200 tweets for validation purposes. Additionally, it includes 400 unlabeled tweets intended for testing the model's generalization capabilities. Each tweet within the dataset is labeled with two categories: 'hope speech' (hs) or 'non-hope speech' (nhs).

Table 1 provides detailed statistics of the dataset used in this study for each sets with each labels for Spanish language., including the distribution of hope speech and non-hope speech across training, development, and test sets. This balanced dataset is crucial for training unbiased models and reflects the diverse nature of language use in social media, pivotal for the accurate detection of hope speech.

Table 1
Dataset Statistics

Data	Classes	Count of Tweets - Spanish
Training	Hope Speech (HS)	700
	Non Hope Speech (NHS)	700
Development	Hope Speech (HS)	100
	Non Hope Speech (NHS)	100
Test	-	400
Total		2,000

To understand the common terms used in hope speech (HS) and non-hope speech (NHS) tweets, word clouds were generated for both categories (Figure 1). These visualizations highlight the most frequent terms, providing insight into the linguistic features that may differentiate HS from NHS tweets. The HS word cloud prominently features terms like "ser," "lgtbi," and "persona," indicating a focus on identity and inclusivity. In contrast, the NHS word cloud also highlights "lgtbi" and "persona" but includes more varied terms reflecting broader discourse.

---

[1]https://codalab.lisn.upsaclay.fr/competitions/17714
[2]https://codalab.lisn.upsaclay.fr/competitions/17714#participate-get_data

Figure 1: Word clouds for HS and NHS tweets

Analyzing the text length distribution helps in understanding the typical length of tweets in each category, which can inform preprocessing steps such as padding or truncation during model training. The text length distribution for the validation set shows that HS tweets are often longer, with a concentration around 275 characters (Figure 3). NHS tweets in the validation set exhibit a wider distribution, with peaks around 150 and 275 characters. The training set exhibits a similar pattern, with HS tweets predominantly longer and NHS tweets showing more variation in length (Figure 2). The test set shows a varied distribution of tweet lengths, with a notable peak around 275 characters, indicating that many tweets are close to the maximum allowed length on Twitter (Figure 4).
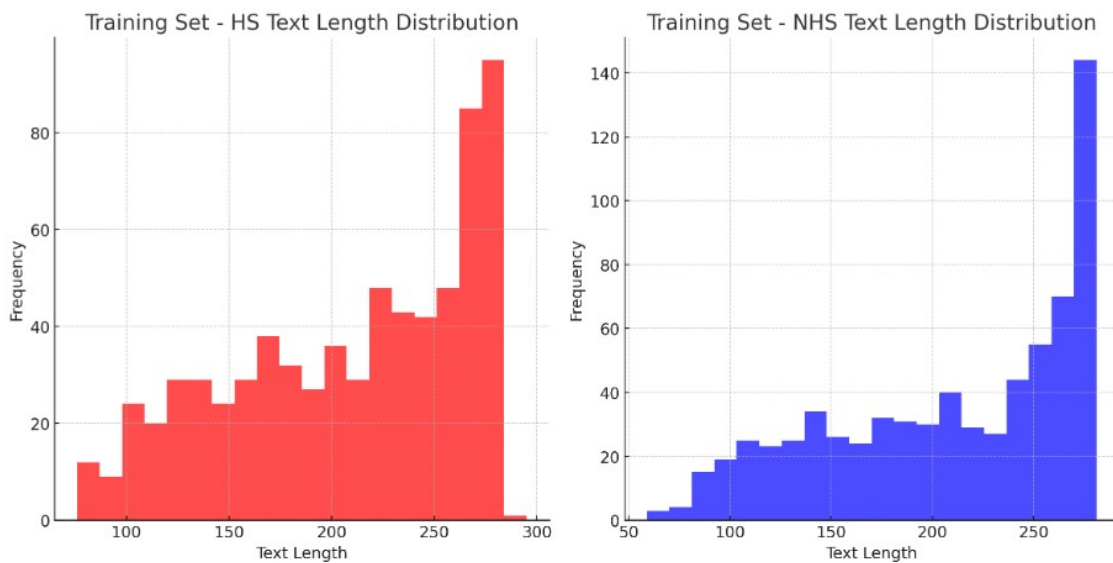

Figure 2: Training Set Length Distribution

These visual analyses provide a comprehensive understanding of the dataset characteristics, aiding in the development of effective NLP models for hope speech detection. From these figures, we can see that the length and content of tweets play a crucial role in distinguishing between hope speech and non-hope speech.

4. Methodology

4.1. Data Preprocessing

The text data preprocessing steps involved removing numerical characters, stripping punctuation, converting text to lowercase, eliminating Spanish stopwords using NLTK, and removing emojis with
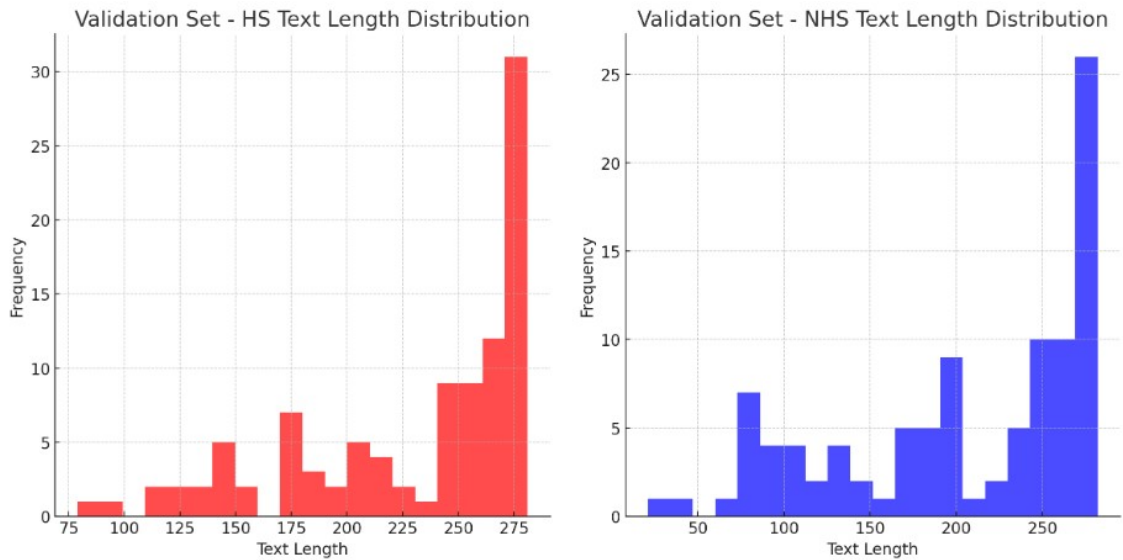
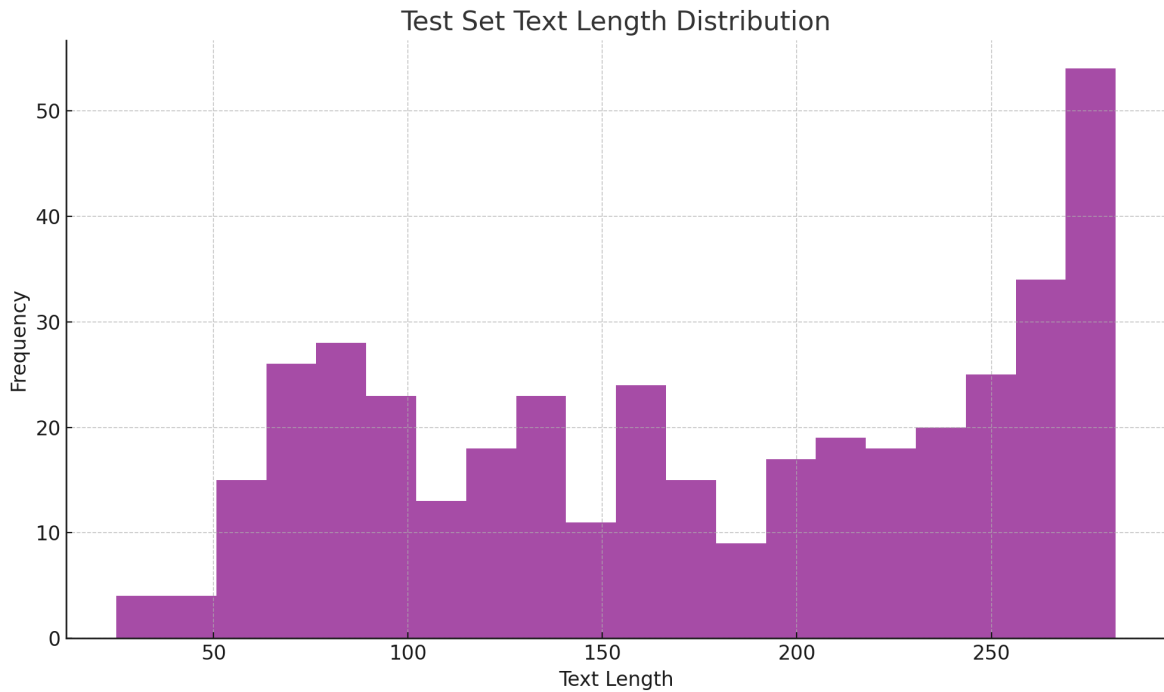**Figure 3:** Validation Set Length Distribution

Figure 4: Test Set Length Distribution

regular expressions. These procedures aimed to enhance the dataset quality for machine learning applications by focusing on linguistic content and meaningful words while maintaining consistency.

4.2. Data Encoding and Batching

Following preprocessing, the text data was tokenized using the AutoTokenizer from the Hugging Face transformers library. The tokenizer converted the text into a sequence of tokens, which were then padded to a consistent length of 128 tokens to ensure uniform input size. These tokenized sequences, along with attention masks indicating real tokens versus padding, and the encoded labels, were encapsulated in a TensorDataset. DataLoaders were subsequently created with a batch size of 16 to facilitate efficient processing during model training.
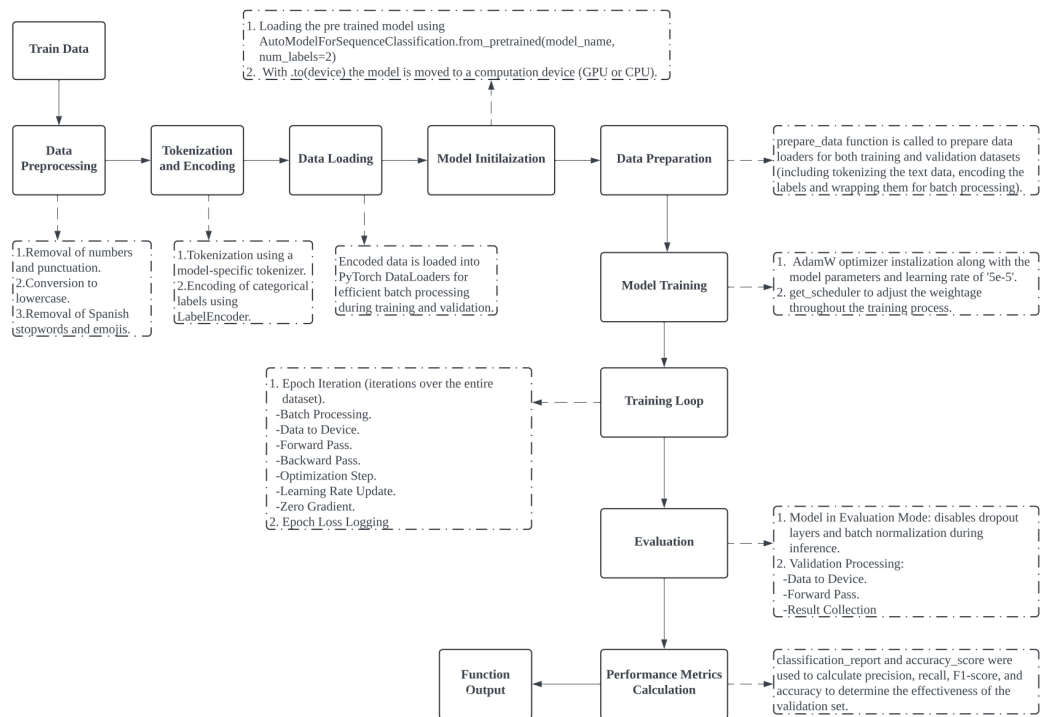
Figure 5: Methodology Workflow

4.3. Model Training and Evaluation

4.3.1. Model Configuration

The models utilized in this study were initialized from pre-trained configurations using the AutoModelForSequenceClassification. This allowed for leveraging the pre-trained knowledge and fine-tuning the models for the specific binary classification task at hand, distinguishing between 'hope speech' and 'non-hope speech'.

4.3.2. Training

The model training phase involved training the models over 10 epochs. The AdamW optimizer was employed with an initial learning rate set at 5×10^{-5}. Additionally, a linear scheduler was utilized to dynamically adjust the learning rate during training based on the model's progress. The training process consisted of forward and backward passes to compute the loss and update the model parameters through gradient descent, optimizing the model for the binary classification task.

4.3.3. Comparative Approaches explored

Following the training phase, the trained models were evaluated on a balanced validation set. Our experiments included a range of models such as bert-base-spanish-wwm-cased-finetuned-spa-squad2-es, bert-base-spanish-wwm-cased, bert-base-spanish-wwm-uncased [23], xlm-roberta-base [24], roberta-base-bne [25] and electricidad-base-discriminator [26]. Moreover, our methodology involved utilizing deep neural network transformer models with contextually sensitive lexical augmentation to enhance the training datasets, thereby generating additional training samples. This approach aligns with current

trends in natural language processing research, emphasizing the significance of data augmentation for enhancing model performance.

Performance metrics such as accuracy and F1-score were computed to assess the models' effectiveness in distinguishing between 'hope speech' and 'non-hope speech'. Furthermore, detailed classification reports were generated to provide insights into the model's performance, highlighting its ability to correctly classify instances of 'hope speech' and 'non-hope speech'. Model performances were visualized using bar charts to facilitate a comparative analysis of different models based on their accuracy and F1 scores. This visual representation aided in identifying the most effective models for hope speech detection.

5. Results

5.1. Overview of Model Performance

The study evaluates the performance of six transformer-based models on the task of detecting hope speech within tweets from the LGBTQI community. Each model was tested on a balanced validation set, labeled as 'hope speech' (hs) or 'non-hope speech' (nhs). The models varied in their approach and specialization in handling Spanish language texts. Below, we present a summary of their performance in terms of precision, recall, F1 score, and overall accuracy.

5.2. Performance Metrics

The performance of the models is presented in Table 2. This table captures detailed metrics including precision, recall, f1-score, and overall accuracy, assessed on a balanced validation set consisting of tweets categorized into 'hope speech' (hs) and 'non-hope speech' (nhs).

5.3. Analysis of Results

5.3.1. Performance Discrepancies Across Models

The dccuchile/bert-base-spanish-wwm-cased model exhibited superior performance with the highest overall accuracy (0.84). This model achieved particularly high precision in detecting hope speech (0.90) and excellent recall in identifying non-hope speech (0.91). The high precision indicates that when the model predicts a tweet as hope speech, it is very likely correct, making it a reliable choice for applications where false positives (non-hope speech incorrectly labeled as hope speech) are particularly problematic. The high recall in non-hope speech suggests that the model is also effective at capturing most instances of non-hope speech, thereby reducing false negatives (hope speech incorrectly labeled as non-hope speech). Conversely, the mrm8488/electricidad-base-discriminator model, while still performing adequately, showed the lowest accuracy among the tested models at (0.76). Despite this, its balanced precision and recall metrics indicate a conservative approach to classification—neither overly penalizing nor overly rewarding any particular class. This balance makes it suitable for scenarios where maintaining a moderate level of detection across categories is more crucial than achieving high performance in one at the expense of the other.

5.3.2. Effectiveness in Linguistic Context Handling

Models like the xlm-roberta-base and PlanTL-GOB-ES/roberta-base-bne performed comparably well, with accuracies of 0.82. These models showed a balanced capability in handling both hope and non-hope speech, which reflects their robustness in linguistic context handling. Their performance underscores the capability of transformer-based models to adapt to the subtleties of language used in different social contexts, which is essential for applications deployed across diverse social media platforms.

Table 2
Model Performance Evaluation

Methods	Metrics	Hope	Non-Hope	Macro Avg	Weighted Avg
dccuchile/bert-base-spanish-wwm-cased	Precision	0.81	**0.90**	0.85	0.84
	Recall	**0.91**	0.78	0.84	**0.84**
	F1-score	0.85	0.83	0.84	**0.84**
dccuchile/bert-base-spanish-wwm-uncased	Precision	0.79	0.85	0.82	0.81
	Recall	0.86	0.77	0.81	0.81
	F1-score	0.82	0.81	0.81	0.81
xlm-roberta-base	Precision	**0.84**	0.82	0.83	0.82
	Recall	0.81	**0.84**	0.82	0.82
	F1-score	0.82	**0.83**	0.82	0.82
PlanTL-GOB-ES/roberta-base-bne	Precision	0.83	0.81	0.82	0.82
	Recall	0.81	0.83	0.82	0.82
	F1-score	0.82	0.82	0.82	0.82
mrm8488/electricidad-base-discriminator	Precision	0.73	0.80	0.76	0.76
	Recall	0.82	0.70	0.76	0.76
	F1-score	0.77	0.74	0.76	0.76
mrm8488/bert-base-spanish-wwm-cased-finetuned-spa-squad2-es	Precision	0.76	0.85	0.81	0.80
	Recall	0.87	0.73	0.80	0.80
	F1-score	0.81	0.78	0.80	0.80

5.3.3. Implications of Model Specificities

The nuanced differences in model performance also shed light on the implications of model architecture and training specifics. For instance, models fine-tuned on Spanish-language corpora or those specifically optimized for question-answering tasks (like the mrm8488/bert-base-spanish-wwm-cased-finetuned-spa-squad2-es) indicate that domain-specific training can influence model behavior significantly, especially in terms of how they interpret the contextual and emotional nuances of a text.

6. Conclusion

This research has demonstrated the efficacy of deep learning models in the detection of hope speech across social media platforms, underscoring their potential to foster positive online environments. Practically, these models can be integrated into social media monitoring tools to automatically flag and promote hopeful content, thereby enhancing user engagement and creating supportive online communities. Additionally, organizations focused on mental health and community support can utilize these models to identify and amplify positive messages, contributing to the well-being of marginalized groups. By employing advanced NLP techniques, the study has effectively identified linguistic patterns indicative of hope speech, which is particularly crucial for supporting marginalized communities. The application of models like BERT and RoBERTa has revealed not only high accuracy in classifying text but also the ability to adapt to the nuanced expressions of hope in different linguistic contexts.

Furthermore, the results highlight the importance of continued development in machine learning to enhance the sensitivity and specificity of these models. This is vital for reducing false positives and ensuring that the detection of hope speech does not inadvertently suppress free expression. The

research also calls attention to the need for comprehensive datasets that reflect the diversity of language use across various demographics and geographies to improve model generalization.

In conclusion, while significant strides have been made, the path forward involves refining these technologies through rigorous testing and broadening their linguistic and cultural scope. This ongoing work will contribute to more inclusive and supportive online communities, ultimately using technology to elevate the quality of social discourse and resilience among users.

6.1. Future Research Directions

Given the observed performance disparities, future research could investigate combining the strengths of these models through ensemble techniques, which could leverage the high precision of one model and the high recall of another to achieve better overall performance. Moreover, exploring additional linguistic features and integrating socio-linguistic context more deeply could further enhance the models' ability to discern hope speech from varied textual inputs.

Acknowledgement

References

[1] R. Vogel, F. Hattke, A century of Public Administration: Traveling through time and topics, Public Administration 100 (2022) 17–40.

[2] S. Fyffe, P. Lee, S. Kaplan, "Transforming" personality scale development: Illustrating the potential of state-of-the-art natural language processing, Organizational Research Methods 27 (2024) 265–300.

[3] F. Balouchzahi, S. Butt, G. Sidorov, A. Gelbukh, CIC@ LT-EDI-ACL2022: Are transformers the only hope? Hope speech detection for Spanish and English comments, in: Proceedings of the second workshop on language technology for equality, diversity and inclusion, 2022, pp. 206–211.

[4] T. Nath, V. K. Singh, V. Gupta, BongHope: An Annotated Corpus for Bengali Hope Speech Detection (2023).

[5] B. R. Chakravarthi, Hope speech detection in YouTube comments, Social Network Analysis and Mining 12 (2022) 75.

[6] D. Khanna, M. Singh, P. Motlicek, IDIAP_TIET@ LT-EDI-ACL2022: Hope Speech Detection in Social Media using Contextualized BERT with Attention Mechanism, in: Proceedings of the Second Workshop on Language Technology for Equality, Diversity and Inclusion, 2022, pp. 321–325.

[7] S. Palakodety, A. R. KhudaBukhsh, J. G. Carbonell, Hope speech detection: A computational analysis of the voice of peace (2019).

[8] 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.

[9] A. Conneau, K. Khandelwal, N. Goyal, V. Chaudhary, G. Wenzek, F. Guzmán, E. Grave, M. Ott, L. Zettlemoyer, V. Stoyanov, Unsupervised cross-lingual representation learning at scale, arXiv preprint arXiv:1911.02116 (2019).

[10] H. Surana, B. Chinagundi, ginius@ lt-edi-acl2022: Aasha: transformers based hope-edi, in: Proceedings of the Second Workshop on Language Technology for Equality, Diversity and Inclusion, 2022, pp. 291–295.

[11] B. R. Chakravarthi, B. Bharathi, J. P. Mccrae, M. Zarrouk, K. Bali, P. Buitelaar, Proceedings of the

second workshop on language technology for equality, diversity and inclusion, in: Proceedings of the Second Workshop on Language Technology for Equality, Diversity and Inclusion, 2022.

[12] Y. Zhu, LPS@ LT-EDI-ACL2022: an ensemble approach about hope speech detection, in: Proceedings of the Second Workshop on Language Technology for Equality, Diversity and Inclusion, 2022, pp. 183–189.

[13] K. Puranik, A. Hande, R. Priyadharshini, S. Thavareesan, B. R. Chakravarthi, Iiitt@ lt-edi-eacl2021-hope speech detection: there is always hope in transformers, arXiv preprint arXiv:2104.09066 (2021).

[14] B. R. Chakravarthi, V. Muralidaran, R. Priyadharshini, S. Cn, J. P. McCrae, M. Á. García, S. M. Jiménez-Zafra, R. Valencia-García, P. Kumaresan, R. Ponnusamy, et al., Overview of the shared task on hope speech detection for equality, diversity, and inclusion, in: Proceedings of the second workshop on language technology for equality, diversity and inclusion, 2022, pp. 378–388.

[15] M. Orken, O. Dina, A. Keylan, T. Tolganay, O. Mohamed, A study of transformer-based end-to-end speech recognition system for kazakh language, Scientific reports 12 (2022) 8337.

[16] L. B. Letaifa, J.-L. Rouas, Transformer model compression for end-to-end speech recognition on mobile devices, in: 2022 30th European Signal Processing Conference (EUSIPCO), IEEE, 2022, pp. 439–443.

[17] B. R. Chakravarthi, Multilingual hope speech detection in English and Dravidian languages, International Journal of Data Science and Analytics 14 (2022) 389–406.

[18] 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.

[19] 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.

[20] S.-M. Laaksonen, J. Haapoja, T. Kinnunen, M. Nelimarkka, R. Pöyhtäri, The datafication of hate: Expectations and challenges in automated hate speech monitoring, Frontiers in big Data 3 (2020) 3.

[21] 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.

[22] 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, 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).

[23] 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.

[24] A. Conneau, K. Khandelwal, N. Goyal, V. Chaudhary, G. Wenzek, F. Guzmán, E. Grave, M. Ott, L. Zettlemoyer, V. Stoyanov, Unsupervised Cross-lingual Representation Learning at Scale, CoRR abs/1911.02116 (2019). URL: http://arxiv.org/abs/1911.02116. arXiv:1911.02116.

[25] A. Gutiérrez-Fandiño, J. Armengol-Estapé, M. Pàmies, J. Llop-Palao, J. Silveira-Ocampo, C. P. Carrino, C. Armentano-Oller, C. R. Penagos, A. Gonzalez-Agirre, M. Villegas, MarIA: Spanish Language Models, Proces. del Leng. Natural 68 (2021) 39–60. URL: https://api.semanticscholar.org/CorpusID:252847802.

[26] M. Romero, Spanish Electra by Manuel Romero, https://huggingface.co/mrm8488/electricidad-base-discriminator/, 2020.