

Enhancing Hope Speech Detection on Twitter Using Machine Learning and Transformer Models

Lemlem Eyob^{1,*}, Tsadkan Yitbarek², Amna Naseeb¹, Grigori Sidorov¹ and Ildar Batyrshin¹

¹Instituto Politécnico Nacional (IPN), Centro de Investigación en Computación (CIC), Mexico City, Mexico

²Maharishi International University, Fairfield, Iowa

Abstract

Hope is a positive mood rooted in the expectation of favorable results in one's life or the world in general, and it is both expressed in the present and the future. We make use of traditional machine learning models and transformer algorithms such as Support Vector Machine (SVM), Random Forest (RF), and a transformer-based BERT model for hope speech detection using an English dataset for binary hope speech detection collected from Twitter, which is provided by HOPE at the IberLEF 2024 share task organizers. Our experiment using the BERT model achieved a macro-average F1-score of 0.85 in the binary classification task, and when compared to the above-mentioned machine learning models, it consistently outperforms them. This study provides valuable insights into addressing hope speech and explores the effectiveness of advanced NLP techniques in promoting positive communication online.

Keywords

Hope, Not Hope, BERT model, Machine learning

1. Introduction

Hope Speech detection is the process of identifying and detecting inspirational talks, comments, and posts filled with positive vibes [1]. Today, social media platforms online are greatly affecting human life, and people can freely express their thoughts on these social networks [2, 3].

Many studies have been conducted to monitor the spread of negativity in modern times by removing vulgar, offensive [4], hatespeech [5] and threatening comments from social media. Nevertheless, there are fewer studies that concentrate on the fact that positivity is important, promoting the fostering of supportive and reassuring content in online forums [6].

NLP researchers are deeply involved in exploring a wide array of linguistic areas due to the exponential growth of online data. Their investigations encompass sentiment analysis [7], hate speech detection, language identification [8], fake news identification [9], recognition of positive emotions [10], and more.

These efforts are directed at deciphering human expression, understanding text sentiment, recognizing offensive language, determining language origins, distinguishing between authentic and deceptive content, and spotting instances of positivity. Additionally, researchers delve into the critical skill of paraphrasing, essential for tasks like summarizing text and translating between languages. In summary, NLP researchers dedicate themselves to unraveling the complexities of language, driving advancements that facilitate improved communication and understanding across various fields and applications.

The main point of the speech is to motivate people who are depressed, lonely, and stressed by the promise, assurance, tips, and help. So, to sum up, the analysis of hope in social media is a necessary tool that can give information about the direction of the goal-directed behaviors that are vital for well-being and that can provide a lot of new and valuable insights into it [1]. Hence, there is a requirement to identify the hope speeches among the social media [6].

IberLEF 2024, September 2024, Valladolid, Spain

*Corresponding author.

✉ lkawo2023@cic.ipn.mx (L. Eyob); tabebe@miu.edu (T. Yitbarek); nasseba23@cic.ipn.mx (A. Naseeb); sidorov@cic.ipn.mx (G. Sidorov); batyr1@cic.ipn.mx (I. Batyrshin)



© 2024 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

2. Related Work

Nowadays, the digital world has fundamentally changed the way people network and socialize [11]. Quite a lot of investigation has been done on the detection of fake news [12] and hate speech [7, 13] from social media data. Lately, NLP researchers are directing their focus to the automatic detection of hope-speech. Examining hope on social media is now considered a must for comprehending well-being and the path toward goal-directed behaviors. Among the techniques and models used for the detection of hope speech are the following.

Firstly, the authors of [14] investigated the impact of psycholinguistic and linguistic features on hope speech detection using a non-complex deep learning algorithm [15]. Additionally, the MUCS team presented three proposed models for the "Hope Speech Detection for Equality, Diversity, and Inclusion-EACL 2021" task, including CoHope-ML, a machine learning voting classifier, CoHope-NN, a deep learning neural network model, and CoHope-TL, a transfer learning-based model [1].

Moreover, [16] describes a study involving a curated analysis that introduced a two-level dataset for hope speech detection in English tweets, marking the first-ever attempt to address hope speech detection with the actual concept of hope as a multiclass classification task. Another approach proposed in [2] utilized the SMOTE technique to resolve data imbalance issues and a 1D Conv-LSTM model for classification.

Furthermore, various machine learning and deep learning approaches were utilized in the hope speech detection shared task at EACL 2021 [17]. Authors in [6] proposed the creation of an English-Kannada Hope speech dataset, KanHope, and in the study [18] present hope speech detection among posts in English and Spanish using support vector machine (SVM). while [19] presented manually annotated datasets for hope speech detection in English, Tamil, and Malayalam. They experimented with multiple machine learning models, including support vector machine (SVM), logistic regression, K-nearest neighbor, decision tree, and logistic neighbors, and proposed a new CNN-based model, which outperformed others with impressive macro F1-scores for each language. Finally, a transformer-based pre-trained BERT model with a rule-based language identification system was described in [20, 21] for detecting hope speech in YouTube comments.

3. Contributions

- Our work underscores the efficiency and potential of transformer models, particularly BERT, in fostering positive online communication by accurately recognizing hope speech.
- We advance the current state of research by demonstrating how transformer models like BERT can improve the quality of online interactions.
- We highlight the promising approach that sets the stage for future studies and practical applications aimed at enhancing online communication.
- By leveraging the advanced capabilities of BERT and similar models, we contribute to creating a more supportive and positive online environment.
- Our research aims to enhance the overall user experience and promote positive social discourse through the effective use of transformer models.

4. Methodology

We started our experiments with machine learning (ML) algorithms like Support Vector Machines (SVM), Random Forest. However, Transformer based BERT model gave superior results to the above mentioned techniques. We implemented the ML algorithms using scikit-learn We use-IDF vectorization Technique to transform the text data into TF-IDF vectors before feeding it into a machine learning model. This processed data can then be used for further analysis or training machine learning models.

Table 1

Distribution of labels for training and evaluation data

	Hope	Not Hope
Training	3104	3088
Evaluation	530	502 tables

4.1. Share Task Discription

Hope is the set of openness of spirit towards the future that is a desire, expectation and wish for something to happen or to be true that is very important in a human’s state of mind, emotions, behavior, and decision-making [16, 22].

The shared task on Hope Speech, which is “Task 2: Hope as Expectations,” has two approaches:

- Subtask 2 . a: Binary Hope Speech Detection from English and Spanish texts
- Subtask 2 . b: Multiclass Hope Speech Detection from English and Spanish texts.

We work specifically on tasks of Binary Hope Speech Detection from English texts. Considering the dataset that included English tweets, the system should be able to recognize its class. Based on the training data, our team will categorize the text into ‘Hope speech’ and ‘Not Hope speech’. We are working on this shared task of Binary Hope Speech Detection for the English dataset. The assessment is done on the basis of Precision, Recall, and F1 scores.

- Hope: tweets that display a mention of hope.
- Not Hope: tweets that cannot be described as having hope, expectation, or desire.

They also offer training, validation, and test datasets, including the golden test dataset to experiment more.

4.2. Dataset Description

This paper used the corpus from [16] provided by the “HOPE at IberLEF 2024” organizers [23], [24] to train and tune the models. The dataset encompassed English and Spanish tweets originating from the first half of 2022, amounting to an aggregate of approximately 100,000 tweets per language.

However, in our submission, we are attempting this shared task of binary hope speech detection for the English tweets dataset [16]. The comments in each data set have been labeled as either ‘Hope’ if they contain hope speech or, if not, as ‘Not hope’. When a comment is given to the proposed system, it will be classified into one of these classes [17]. The distribution of labels for training and evaluation data is shown in Table 1, and Figure 1 and 2 also show the training and evaluation dataset label distribution.

4.3. Data Preprocessing

Since the comments in raw format are highly unstructured, containing irrelevant information that may cause any AI-based model to malfunction [25]. The dataset was cleaned up and pre-processed before model implementation. The primary tool for preprocessing is the ‘re’ module from Python’s standard library, utilized for working with regular expressions to perform text manipulation tasks. The following steps were employed:-

1. Remove URLs: In this step, any sequence of characters starting with "HTTP" (HTTP), followed by any non-whitespace characters (\$+), is replaced with an empty string (“”), effectively removing URLs from the text.
2. Remove numbers: This step removes all numbers from the text.
3. Remove special characters: This step removes all special characters from the text except whitespace.
4. Remove emojis: This part removes emojis by first encoding the text into ASCII using encode(ASCII, ‘ignore’). Emojis are non-ASCII characters, so this effectively removes them. Then, it decodes the text back to Unicode using decode(‘ascii’).

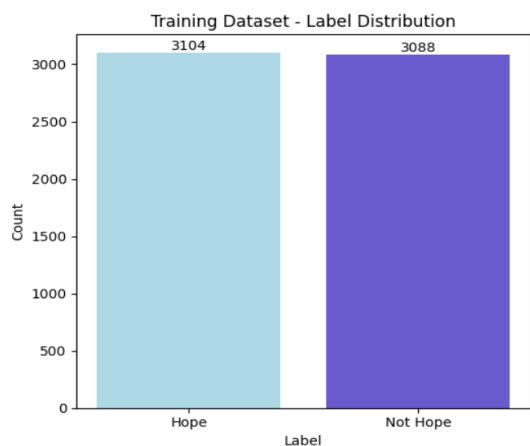


Figure 1: Training dataset label distribution

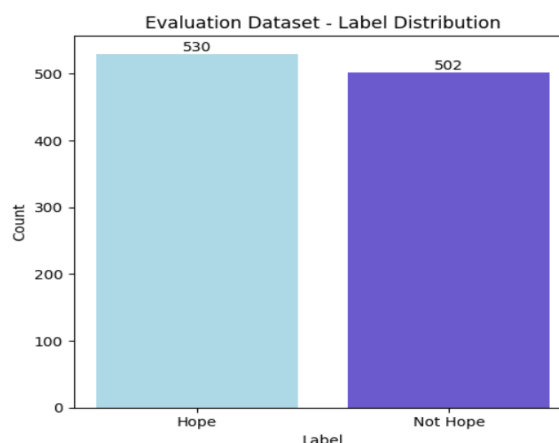


Figure 2: Evaluation dataset label distribution

5. Convert text to lowercase: Finally, the lower() method is used to convert all characters in the text to lowercase.

So, in summary, this function takes a piece of text as input and performs several preprocessing steps to clean it up, including removing URLs, numbers, special characters, and emojis, and converting the text to lowercase. This cleaning process is essential to improving the quality of the data being used. Figure 3 and 4 below show samples of the dataset before and after preprocessing.

	id	text	binary
0	5820	#USER# #USER# I'm really liking this project, ...	Hope
1	4061	#USER# Oh shit really? I would hope they'd she...	Hope
2	1621	#USER# Good morning, Bud! 🍷 Another good decis...	Hope
3	1754	i aspire to have the level of delusion to beli...	Hope
4	401	#USER# #USER# Projects are continuously attack...	Not Hope
...
6187	3889	☀️: But, it was a great depiction of the main c...	Not Hope
6188	3945	Because it's brave to outright misappropriate ...	Not Hope
6189	6736	#USER# They got this one wrong. I'm a devout, ...	Not Hope
6190	4845	#USER# #USER# Not even close. Myself as many o...	Not Hope
6191	4753	#USER# 🍷 agree mate. Started doing this a few ...	Not Hope

6192 rows x 3 columns

Figure 3: Training dataset before preprocessing

	id	text	binary
0	5820	user user im really liking this project lets w...	Hope
1	4061	user oh shit really i would hope theyd shed so...	Hope
2	1621	user good morning bud another good decision f...	Hope
3	1754	i aspire to have the level of delusion to beli...	Hope
4	401	user user projects are continuously attacked b...	Not Hope
...
6187	3889	but it was a great depiction of the main char...	Not Hope
6188	3945	because its brave to outright misappropriate f...	Not Hope
6189	6736	user they got this one wrong im a devout pract...	Not Hope
6190	4845	user user not even close myself as many others...	Not Hope
6191	4753	user agree mate started doing this a few mont...	Not Hope

6192 rows x 3 columns

Figure 4: Training dataset after preprocessing

4.4. Machine Learning Models

We implemented traditional machine learning algorithms, including Random Forest (RF) and Support Vector Machine (SVM), to classify text data for binary hope speech detection.

Random Forest (RF)

Random Forest is an ensemble learning method used for classification and regression tasks, operating by aggregating the results of multiple individual decision trees. In our experiments, the Random Forest classifier demonstrated moderate performance. The model achieved an overall accuracy of 51% and a macro-average F1-score of 0.51. These metrics indicate that while the model is somewhat effective, there is considerable room for improvement.

Support Vector Machine (SVM)

Support Vector Machine is a robust classification algorithm that finds the optimal boundary to separate different classes in the data, maximizing the margin between them. It is particularly effective in high-dimensional spaces, making it suitable for text classification tasks. The SVM model in our study achieved an accuracy of 50% and a macro-average F1-score of 0.50 on the English dataset for binary hope speech detection. This performance is slightly lower than the Random Forest model, further optimization and fine-tuning of the model parameters may be necessary to improve its performance.

4.4.1. Experimental Setups for Machine Learning Models

Feature engineering is a critical step that transforms raw text data into meaningful numerical representations, enabling machine learning models to learn and make accurate predictions. Techniques like TF-IDF, word embeddings, and feature selection help capture the essence of the text while scaling and normalization ensure numerical stability. For advanced models, handling text sequences appropriately is essential. By carefully engineering features, you set a solid foundation for training effective machine learning models. For both of the above mentioned ML models we implemented, We use TF-IDF vectorization for feature extraction.

• Model Training

Two different models are trained:

- `Random Forest Classifier`: Trained with 200 decision trees on the training data.
- `Support Vector Classifier`: Trained with a linear kernel on the TF-IDF transformed training data.

• Text Data

Each entry is extensively annotated for Not Hope or Hope and we provided with separate training, tasting and validation sets. Both experiments assume `X_train` and `y_train` are preprocessed text data and corresponding labels for training, and `dftest['text']` is the text data for testing.

4.5. BERT

Bidirectional Encoder Representations from Transformers (BERT) [26]. BERT model effectively eliminates the number of parameters without affecting performance. The model was chosen because of its outstanding outcomes on other tasks [27]. In this experiment also demonstrated better compared to other traditional Machine Learning models scores with an F1-score of 0.85. Table 2 shows the Parameter values for BERT.

4.5.1. Experimental Setups for BERT

Word embeddings capture the semantic meaning and context of words by representing them as dense vectors in a continuous vector space. BERT generates embeddings using bidirectional context, i.e., analyzes context from both the left and right of a word. Also, BERT's attention architecture computes the attention parallelly for the whole input at once.

We implement and fine-tune a BERT model for binary classification using the 'ktrain' library. It involves importing necessary libraries, initializing a BERT model (`google/bert_uncased_L-12_H-768_A-12`), and setting up the transformer with a maximum input length of 400 tokens. The training and validation data are preprocessed to fit the BERT model requirements. A classifier is then created, and a 'Learner' object is initialized to facilitate training with a batch size of 12 and a number of epochs of 3. Finally, a 'Predictor' object is created to make predictions on the test set, leveraging the pre-trained BERT model's knowledge and fine-tuning it for the specific dataset.

Table 2
Parameters values

Parameter	Value
vocab_size	3000
embedding_dim	100
max_length	200
padding_type	'post'
trunc_type	'post'
num_epochs	10

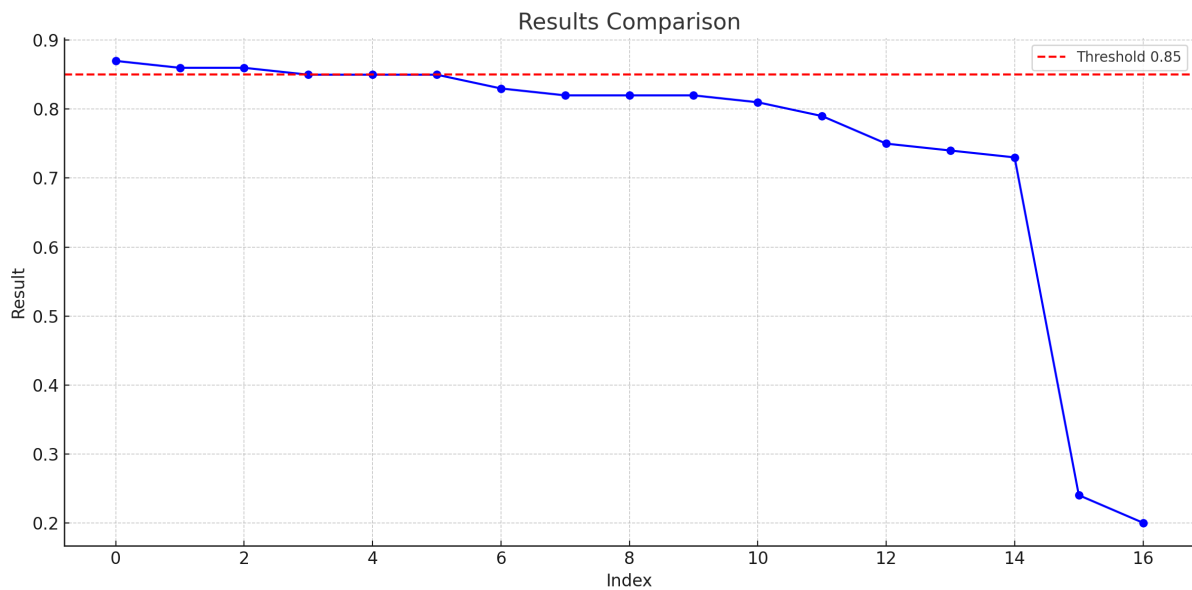


Figure 5: Comparison of the runs submitted

5. Results and Discussion

For this approach, we used models, namely support vector machines (SVM), Random Forest (RF), and a transformer-based BERT mode for English. Out of all these models, the BERT model produced the best results out of all the models. A pre-trained BERT model proved to be the best by yielding a macro-average F1-score of 0.85. The results for all models are shown in Table 3.

Table 3
Evaluation metrics

Model	Precision	Recall	F1-Score	Accuracy
BERT	0.85	0.85	0.85	0.85
Random Forest	0.51	0.51	0.51	0.51
Support Vector Machine	0.50	0.50	0.50	0.50

The line graph 5 shows the results of the participants. The red dashed line marks the threshold of 0.85 for comparison. The graph compares the performance results of various participants, with a specific threshold set at 0.85. The result of the BERT model in our experiment, which is exactly 0.85, is highlighted against the results of other participants. The graph shows that the result in our experiment is above average and meets the threshold, placing us in the upper range of the performance spectrum compared to others. This indicates a strong performance relative to the group.

6. Conclusion

In our study, we tackled the complex task of detecting hope speech using advanced machine learning techniques, with a particular focus on the BERT Transformer approach applied to an English dataset provided by the HOPE at IberLEF 2024 shared task organizers. We compared the performance of traditional machine learning methods, such as Support Vector Machine (SVM) and Random Forest, against the state-of-the-art BERT model. Our findings revealed that the BERT model significantly outperformed these traditional methods, achieving an impressive F1-score of 0.85, thus demonstrating its superior natural language processing capabilities in identifying hope speech.

Our experiments also highlighted the importance of TF-IDF vectorization in the preprocessing of text data, which was crucial for the effectiveness of the machine learning models. We found that the TF-IDF vectorization technique provided a robust foundation for feature extraction, enabling the models to better understand and classify the text data.

Looking forward, we plan to expand our research by incorporating larger datasets, which we believe will provide a more comprehensive understanding and enable us to fine-tune the models further. Further optimization and fine-tuning of the model parameters are necessary. Potential steps include hyperparameter tuning and exploring different feature engineering techniques. By doing so, we aim to achieve higher accuracy and better overall performance in future iterations.

Acknowledgments

The work was done with partial support from the Mexican Government through the grant A1-S-47854 of CONACYT, Mexico, and grants 20241816, 20241819, and 20240951 of the Secretaría de Investigación y Posgrado of the Instituto Politécnico Nacional, Mexico. The authors thank the CONACYT for the computing resources brought to them through the Plataforma de Aprendizaje Profundo para Tecnologías del Lenguaje of the Laboratorio de Supercómputo of the INAOE, Mexico, and acknowledge the support of Microsoft through Microsoft Latin America PhD Award.

References

- [1] F. Balouchzahi, B. Aparna, H. Shashirekha, Mucs@ It-edi-eacl2021: Cohope-hope speech detection for equality, diversity, and inclusion in code-mixed texts, in: Proceedings of the First Workshop on Language Technology for Equality, Diversity and Inclusion, 2021, pp. 180–187.
- [2] A. Gowda, F. Balouchzahi, H. Shashirekha, G. Sidorov, Mucic@ It-edi-acl2022: Hope speech detection using data re-sampling and 1d conv-lstm, in: Proceedings of the second workshop on language technology for equality, diversity and inclusion, 2022, pp. 161–166.
- [3] G. Bade, O. Kolesnikova, G. Sidorov, J. Oropeza, Social media fake news classification using machine learning algorithm, in: Proceedings of the Fourth Workshop on Speech, Vision, and Language Technologies for Dravidian Languages, 2024, pp. 24–29.
- [4] M. Zamir, M. Tash, Z. Ahani, A. Gelbukh, G. Sidorov, Lidoma@ dravidianlangtech 2024: Identifying hate speech in telugu code-mixed: A bert multilingual, in: Proceedings of the Fourth Workshop on Speech, Vision, and Language Technologies for Dravidian Languages, 2024, pp. 101–106.
- [5] M. Shahiki-Tash, J. Armenta-Segura, Z. Ahani, O. Kolesnikova, G. Sidorov, A. Gelbukh, Lidoma at homomex2023@ iberlef: Hate speech detection towards the mexican spanish-speaking lgbt+ population. the importance of preprocessing before using bert-based models, in: Proceedings of the Iberian Languages Evaluation Forum (IberLEF 2023), 2023.
- [6] A. Hande, R. Priyadharshini, A. Sampath, K. P. Thamburaj, P. Chandran, B. R. Chakravarthi, Hope speech detection in under-resourced kannada language, arXiv preprint arXiv:2108.04616 (2021).
- [7] M. G. Yigezu, T. Kebede, O. Kolesnikova, G. Sidorov, A. Gelbukh, Habesha@ dravidianlangtech: Utilizing deep and transfer learning approaches for sentiment analysis., in: Proceedings of the Third Workshop on Speech and Language Technologies for Dravidian Languages, 2023, pp. 239–243.

- [8] M. S. Tash, Z. Ahani, A. Tonja, M. Gameda, N. Hussain, O. Kolesnikova, Word level language identification in code-mixed kannada-english texts using traditional machine learning algorithms, in: Proceedings of the 19th International Conference on Natural Language Processing (ICON): Shared Task on Word Level Language Identification in Code-mixed Kannada-English Texts, 2022, pp. 25–28.
- [9] M. Zamir, M. Tash, Z. Ahani, A. Gelbukh, G. Sidorov, Tayyab@ dravidianlangtech 2024: detecting fake news in malayalam lstm approach and challenges, in: Proceedings of the Fourth Workshop on Speech, Vision, and Language Technologies for Dravidian Languages, 2024, pp. 113–118.
- [10] M. S. Tash, Z. Ahani, O. Kolesnikova, G. Sidorov, Analyzing emotional trends from x platform using senticnet: A comparative analysis with cryptocurrency price, arXiv preprint arXiv:2405.03084 (2024).
- [11] A. L. Tonja, M. G. Yigezu, O. Kolesnikova, M. S. Tash, G. Sidorov, A. Gelbukh, Transformer-based model for word level language identification in code-mixed kannada-english texts, arXiv preprint arXiv:2211.14459 (2022).
- [12] M. Yigezu, O. Kolesnikova, G. Sidorov, A. Gelbukh, Habesha@ dravidianlangtech 2024: Detecting fake news detection in dravidian languages using deep learning, in: Proceedings of the Fourth Workshop on Speech, Vision, and Language Technologies for Dravidian Languages, 2024, pp. 156–161.
- [13] Z. Ahani, M. Tash, M. Zamir, I. Gelbukh, Zavira@ dravidianlangtech 2024: Telugu hate speech detection using lstm, in: Proceedings of the Fourth Workshop on Speech, Vision, and Language Technologies for Dravidian Languages, 2024, pp. 107–112.
- [14] F. Balouchzahi, S. Butt, G. Sidorov, A. Gelbukh, Cic@ It-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.
- [15] Z. Ahani, M. Shahiki Tash, Y. Ledo Mezquita, J. Angel, Utilizing deep learning models for the identification of enhancers and super-enhancers based on genomic and epigenomic features, Journal of Intelligent & Fuzzy Systems (2024) 1–11.
- [16] 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.
- [17] M. D. S. S. Eswar, N. Balaji, V. S. Sarma, Y. C. Krishna, S. Thara, Hope speech detection in tamil and english language, in: 2022 International Conference on Inventive Computation Technologies (ICICT), IEEE, 2022, pp. 51–56.
- [18] M. G. Yigezu, G. Y. Bade, O. Kolesnikova, G. Sidorov, A. Gelbukh, Multilingual hope speech detection using machine learning (2023).
- [19] B. R. Chakravarthi, Hope speech detection in youtube comments, Social Network Analysis and Mining 12 (2022) 75.
- [20] S. Gundapu, R. Mamidi, Autobots@ It-edi-eacl2021: one world, one family: hope speech detection with bert transformer model, in: Proceedings of the First Workshop on Language Technology for Equality, Diversity and Inclusion, 2021, pp. 143–148.
- [21] 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.
- [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, 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).
- [23] 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.
- [24] 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

(2023) 1–28.

- [25] 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.
- [26] J. Devlin, M.-W. Chang, K. Lee, K. Toutanova, Bert: Pre-training of deep bidirectional transformers for language understanding, *arXiv preprint arXiv:1810.04805* (2018).
- [27] F. Ullah, M. Zamir, M. Arif, M. Ahmad, E. Felipe-Riveron, A. Gelbukh, `Fida@` `dravidianlangtech 2024`: A novel approach to hate speech detection using distilbert-base-multilingual-cased, in: *Proceedings of the Fourth Workshop on Speech, Vision, and Language Technologies for Dravidian Languages, 2024*, pp. 85–90.