# Parameter-Free Spanish Hope Speech Detection

Michael Ibrahim

*Computer Engineering Department, Cairo University, 1 Gamaa Street, 12613, Giza, Egypt*

**Abstract**

Deep learning has gained significant traction in text classification and other natural language processing tasks in recent years. Deep neural networks normally require substantial training, extensive datasets, and resource-consuming hyperparameter tuning. A novel low-resource approach has emerged as a substitute for deep learning methods. Though simple, using Gzip compression for text classification performs unexpectedly well compared to complex models like BERT in various tasks. In this study, the proposed method effectively identified Spanish hope speech. The proposed method was ranked third in the IberLEF 2024 Task on Hope Speech Detection subtask 1 achieving an average Macro $F_1$ score of 0.6522.

**Keywords**

Text Classification, Hope Speech Detection, Compression, Machine Learning, K-Nearest-Neighbors

## 1. Introduction

Billions of internet users can now voice their thoughts and share perspectives through social media platforms. This experience has resulted in both positive and negative exchange of ideas, the former being hope speech and the latter being hate speech [1]. Usually, social media comments and posts are analyzed using techniques like hate speech detection [2], offensive language identification[3], and abusive language detection[4] to limit the spread of negativity.

Lately, hope speeches have gained significant prominence for their ability to to soothe hostility [5] and provide encouragement, suggestions, and inspiration [6] during illness, stress, loneliness, or depression. Social media serves as a platform for offensive messages aimed at individuals based on their race, color, ethnicity, gender, sexual orientation, nationality, or religion. According to [6], social media significantly impacts the persona and society's perception of vulnerable individuals [7, 8, 9]. The LGBT community, racial minorities, and people with disabilities are among the vulnerable groups.

IberLEF 2024 [10] Task - HOPE - subtask 1 [11] - emphasizes detecting hope speech to promote Equality, Diversity, and Inclusion and to mitigate hostility and support individuals facing challenges like illness, stress, or loneliness, particularly within vulnerable groups such as the LGBT community and racial minorities. This task aims to classify Spanish tweets, given a Spanish tweet, the task is to determine whether the tweet conveys hope speech. Specifically, this task is divided into two subtasks: 1.a: Hope speech detection on LGTBI domain; and subtask 1.b: Hope speech detection on unknown domains.

The rest of the paper is organized as follows. The related work is summarized in Section 2. The dataset used for training and validation was detailed in Section 3. In Section 4, the system is presented. Section 5 summarizes the study's key findings.

## 2. Related Work

Researchers were primarily drawn to studying the speech due to its provision of a wide range of information, including emotional states [12]. Literature has extensively analyzed speeches to identify

*IberLEF 2024, September 2024, Valladolid, Spain*

michael.nawar@eng.cu.edu.eg (M. Ibrahim)
www.linkedin.com/in/michael-ibrahim-90 (M. Ibrahim)
0000-0003-2340-8917 (M. Ibrahim)

CEUR Workshop Proceedings ceur-ws.org ISSN 1613-0073

both negative and positive effects, including identifying abusive language that causes violence and encouraging comments that provoke assurance. This study focuses on analyzing hope speech.

In [13], a multilingual dataset was compiled from comments on YouTube in different languages and experimented with two kinds of statistical models to identify positivity One involves machine learning models, and the other deep neural networks. The CNN model, as proposed by the authors, outperformed other models. Saumya and Mishra [14] presented a solution for a system dealing with a comparable challenge [15]. Using YouTube and multiple machine learning, deep learning, and hybrid models in common, they detected hope speech in English, Tamil, and Malay. Deep learning models exceed conventional machine learning models in most experiments. The LSTM and BiLSTM hybrid learning model gave the best results.

In [16], the authors proposed a customized version of several transformation-based pre-trained models, the authors implemented changes such as freezing the model, modifying loss functions, and redefining the final layer. In English test datasets, pre-trained models outperform other methods without customization. In [17], the authors recorded English tweets from Twitter for their study. The dataset was tested against three distinct baselines: traditional machine learning, deep learning, and transformers. The authors have differing perspectives on whether CNN and BiLSTM architectures are considered deep learning or transformer models, despite both being part of the Deep Learning family. The annotation process for classifying the dataset and its benchmarking experiments were presented in detail.

A recently proposed text classification method utilizing Gzip compression outperforms conventional deep learning architectures like transformers [18]. The method leverages lossless compressors, like Gzip, to determine patterns statistically and assign shorter codes to common sequences. Texts in the same category have similar regularity and, as a result, are close in compression space when measured using the normalized compression distance (NCD) metric[19]. Under the NCD metric, a k-nearest neighbor classifier is employed for text classification.

## 3. Data

The SpanishHopeEDI [20] dataset consists of 1,650 LGBT-related tweets annotated as HS (Hope Speech) or NHS (Non-Hope Speech) was used for training and validating the proposed system. In this dataset, a tweet is marked as HS if it supports the social integration of minorities, inspires positivity within the LGTBI community, encourages LGTBI individuals in trying situations, and unconditionally promotes tolerance. A tweet is marked as NHS only if it contains negative sentiment towards the LGTBI community, explicitly seeks violence, or uses gender-based insults. The collection and annotation process is described below.

A tweet is marked as HS if the text:

- advocates for the inclusion of minorities in society.
- serves as a positive inspiration for the LGBT community.
- explicitly encourages individuals identifying as LGBT in potentially challenging situations.
- or, wholeheartedly endorses tolerance.

A tweet is marked as NHS if the text:

- shows no support for the LGBT community.
- advocates for violence.
- or, employs insults based on gender

The agreement among annotators was measured by CohenâĂŹs Kappa[21] and KrippendorfâĂŹs Alpha [22], yielding results of 0.87 and 88.1% respectively, confirming the corpusâĂŹs high-quality.

For the purpose of this subtask [11], a subset of the SpanishHopeEDI [20] dataset was used for training and validating the system, 700 HS and 700 NHS labeled tweets were utilized for training, and 100 HS and 100 NHS tweets for validation. The final test set comprised 400 tweets, 200 of which were HS and the other 200 were NHS.

# 4. Methodology

The first step of a classification model is to clean the dataset. The following algorithm 1 was used to clean up text by deleting non-alphanumeric symbols, deleting URLs, deleting HTML tags, deleting punctuations, deleting Spanish stop words, and deleting emojis.

Code Listing 1: Python Code for Text Normalization

```
from stop_words import get_stop_words
import re

sw = get_stop_words('Spanish)
def clean_text(text):
    text = re.sub(r"[^a-zA-Z?.!,Â£]+", " ", text) # replacing everything with space except (a-z, A
                                                  -Z, ".", "?", "!", ",")

    text = re.sub(r"http\S+", "",text) #Removing URLs

    html=re.compile(r'<.*?>')

    text = html.sub(r'',text) #Removing html tags

    punctuations = '@#!?+&*[]-%.:/();$=><|{}^' + "'`" + '_'
    for p in punctuations:
        text = text.replace(p,'') #Removing punctuations

    text = [word.lower() for word in text.split() if word.lower() not in sw]

    text = " ".join(text) #removing stopwords

    emoji_pattern = re.compile("["
                               u"\U0001F600-\U0001F64F"  # emoticons
                               u"\U0001F300-\U0001F5FF"  # symbols & pictographs
                               u"\U0001F680-\U0001F6FF"  # transport & map symbols
                               u"\U0001F1E0-\U0001F1FF"  # flags (iOS)
                               u"\U00002702-\U000027B0"
                               u"\U000024C2-\U0001F251"
                               "]+", flags=re.UNICODE)
    text = emoji_pattern.sub(r'', text) #Removing emojis

    return text
```

The next step is to apply the compression-based text classification method 2. In this method, the similarity between text instances was measured using Normalized Compression Distance (NCD), resulting in distance matrices for each pair within the training and test sets. Based on two strings, their NCD is determined as follows:

$$NCD(x_1, x_2) = \frac{C(x_1x_2) - \min[C(x_1), C(x_2)]}{\max[C(x_1), C(x_2)]}$$

where $C(x)$ is the size of a compressed string $x$, and this measurement is the basis for training a classifier, and NCD is a key feature of this classification system.

From cross-validation, it was found that the KNN classifier, with a k value of 11, achieved the highest Macro $F_1$ score. and when applied to the test set, the 11-NN classifier achieved an $F_1$ score of 0.6602 for hope speech detection on LGTBI domain; and and $F_1$ score of 0.6442 for hope speech detection on unknown domains, which yields an average Macro $F_1$ score of 0.6522.

Code Listing 2: Python Code for Text Classification with Gzip

```
import gzip
import numpy as np
```

```
for ( x1 , _ ) in test_set :
  Cx1 = len( gzip . compress ( x1.encode () ))
  distance_from_x1 = []
  for ( x2 , _ ) in training_set :
    Cx2 = len( gzip . compress ( x2 .encode () )
    x1x2 = " ". join ([ x1 , x2 ])
    Cx1x2 = len( gzip . compress ( x1x2 .encode () )
    ncd = ( Cx1x2 - min( Cx1 , Cx2 ) ) / max( Cx1 , Cx2 )
    distance_from_x1 . append ( ncd )
  sorted_idx = np . argsort ( np.array (distance_from_x1 ) )
  top_k_class = training_set [sorted_idx [: k ] , 1]
  predict_class = max(set( top_k_class ), key = top_k_class . count )
```

## 5. Conclusion

As online content grows considerably, it is necessary to promote positivity, such as in the form of a hopeful speech on an online forum, to encourage empathy and acceptable social behavior. This paper presented a low-resource approach using Gzip compression for Spanish hope speech detection. The results were promising, in identifying Spanish hope speech, achieving an $F_1$ score of 0.6602 for hope speech detection on LGTBI domain; and and $F_1$ score of 0.6442 for hope speech detection on unknown domains, which yields an average Macro $F_1$ score of 0.6522.

## References

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

[2] A. Schmidt, M. Wiegand, A survey on hate speech detection using natural language processing, in: Proceedings of the fifth international workshop on natural language processing for social media, 2017, pp. 1–10.

[3] H. Yenala, A. Jhanwar, M. K. Chinnakotla, J. Goyal, Deep learning for detecting inappropriate content in text, International Journal of Data Science and Analytics 6 (2018) 273–286.

[4] Y. Lee, S. Yoon, K. Jung, Comparative studies of detecting abusive language on twitter, arXiv preprint arXiv:1808.10245 (2018).

[5] S. Palakodety, A. R. KhudaBukhsh, J. G. Carbonell, Hope speech detection: A computational analysis of the voice of peace, arXiv preprint arXiv:1909.12940 (2019).

[6] B. R. Chakravarthi, Hopeedi: A multilingual hope speech detection dataset for equality, diversity, and inclusion, in: Proceedings of the Third Workshop on Computational Modeling of People's Opinions, Personality, and Emotion's in Social Media, 2020, pp. 41–53.

[7] P. Burnap, W. Colombo, J. Scourfield, Machine classification and analysis of suicide-related communication on twitter, in: Proceedings of the 26th ACM conference on hypertext & social media, 2015, pp. 75–84.

[8] V. Kitzie, " i pretended to be a boy on the internet": Navigating affordances and constraints of social networking sites and search engines for lgbtq+ identity work, First Monday (2018).

[9] D. N. Milne, G. Pink, B. Hachey, R. A. Calvo, Clpsych 2016 shared task: Triaging content in online peer-support forums, in: Proceedings of the third workshop on computational linguistics and clinical psychology, 2016, pp. 118–127.

[10] 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.

[11] 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).

[12] H. Nourtel, P. Champion, D. Jouvet, A. Larcher, M. Tahon, Evaluation of speaker anonymization on emotional speech, arXiv preprint arXiv:2305.01759 (2023).

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

[14] 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).

[15] 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.

[16] N. Ghanghor, R. Ponnusamy, P. K. Kumaresan, R. Priyadharshini, S. Thavareesan, B. R. Chakravarthi, Iiitk@ lt-edi-eacl2021: Hope speech detection for equality, diversity, and inclusion in tamil, malayalam and english, in: Proceedings of the First Workshop on Language Technology for Equality, Diversity and Inclusion, 2021, pp. 197–203.

[17] F. Balouchzahi, G. Sidorov, A. Gelbukh, Polyhope: Two-level hope speech detection from tweets, Expert Systems with Applications 225 (2023) 120078.

[18] Z. Jiang, M. Y. Yang, M. Tsirlin, R. Tang, J. Lin, Less is more: Parameter-free text classification with gzip, arXiv preprint arXiv:2212.09410 (2022).

[19] M. Li, X. Chen, X. Li, B. Ma, P. M. Vitányi, The similarity metric, IEEE transactions on Information Theory 50 (2004) 3250–3264.

[20] 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.

[21] J. Cohen, A coefficient of agreement for nominal scales, Educational and psychological measurement 20 (1960) 37–46.

[22] K. Krippendorff, Agreement and information in the reliability of coding, Communication methods and measures 5 (2011) 93–112.