Hope on the Horizon: Experiments with Learning Models for Hope Speech Detection in Spanish and English

Sonith Divakaran*,†, Kavya Girish† and Hosahalli Lakshmaiah Shashirekha†

Department of Computer Science, Mangalore University, Mangalore, Karnataka, India

Abstract

Hope is the expectation or belief that results in positive outcomes despite challenges or adversity and "hope speech" refers to textual content aimed at inspiring hope, motivation, or positivity. Individuals who are experiencing challenges may be inspired from hope speech as it instills optimism, encouragement, and positivity. On social media platforms, hope speech contributes to fostering a positive and supportive community, offering comfort, motivation, and solidarity to online users. A healthy social media echo system can be maintained by identifying and amplifying the hope speech content. However, identifying hope speech content manually is challenging due to ever growing social media and its users. In this direction, "HOPE at IberLEF 2024" shared task organized at IberLEF 2024, invites the research community to address the challenges of detecting hope speech in Spanish and English languages. The shared task consists of two tasks: Task 1: Hope for Equality, Diversity and Inclusion (HopeEDI) - a binary classification problem in Spanish language, and Task 2: Hope as Expectations, consisting of two subtasks: 2.a) Binary Hope Speech detection and 2.b) Multiclass Hope Speech detection, both in Spanish and English languages. To explore the strategies for detecting hope speech in Spanish and English on social media platforms, in this paper, we - team MUCS, describe the models submitted to these tasks. Various Machine Learning (ML) models and Transfer Learning (TL) techniques are proposed to classify the given Spanish and English text into : i) one of the two categories - 'Not Hope' or 'Hope' in case of binary classification and ii) one of the four categories - 'Not Hope', 'Generalized hope', 'Unrealistic Hope', or 'Realistic Hope', in case of multiclass classification. While Term Frequency-Inverse Document Frequency (TF-IDF) of word n-grams in the range (1, 3), multilingual embeddings, and aligned word vectors, are used as features to train the different ML models individually, Bidirectional Encoder Representations from Transformers (BERT) variants are fine-tuned in TL, to detect hope speech in Spanish and English. The best results obtained by our proposed models are: 10^{th} rank with 0.59 macro F1 score in Task 1 HopeEDI, 5^{th} and 9^{th} ranks with macro F1 scores 0.82 and 0.82 in subtask 2.a) Binary Hope Speech detection in Spanish and English respectively, and 4^{th} and 8^{th} ranks with macro F1 scores 0.64 and 0.56 in subtask 2.b) Multiclass Hope Speech detection in Spanish and English respectively.

Keywords

Hope Speech, Machine Learning, Transfer Learning, Multilingual Models, Transformers

1. Introduction

Social media has fundamentally altered the way people communicate in their daily life to express their ideas/thoughts towards the other person, a community or even an event. As users leverage social media platforms to raise awareness about various social issues and support the causes, these platforms have also emerged as a powerful tool for various types of activities [1]. While some users are misusing the anonymity of users on social media platforms to spread abusive and offensive content, targeting an individual or a group, some good samaritans are spreading motivating or hope speech content to bring a sense of hope and encouragement to the online users who are in distress [2]. Hope content comprises of expressions, tales, and messages that evoke hope, desire, and ignite a sense of motivation among the online users. Additionally, hope speech provides important insights into the values, goals, and general mindset of the society. In general, any positive aspects of social media messages fall under the category of hope speech [2, 3]. Further, hope speech can be categorized as generalistic hope, realistic hope, and

IberLEF 2024, September 2024, Valladolid, Spain

*Corresponding author.

†These authors contributed equally.

sonithksd@gmail.com (S. Divakaran); kavyamujk@gmail.com (K. Girish); hlsrekha@mangaloreuniversity.ac.in (H. L. Shashirekha)

0000-0002-0877-7063 (S. Divakaran); 0000-0001-7116-9338 (K. Girish); 0000-0002-9421-8566 (H. L. Shashirekha)

Table 1
Task description and statistics of the datasets

Task Name	Language	Subtask	Category	Train set	Val set	Test set
Task 1: Hope for Equality, Diversity and Inclusion	Spanish	Hope Speech detection (Train set consists of text in LGBTI domain and test set in unknown domains)	nhs	700	100	200
			hs	700	100	200
Task 2: Hope as Expectations	Spanish	2.a) Binary Hope Speech detection	Not Hope	4,701	799	773
			Hope	2,202	351	379
		2. b) Multiclass Hope Speech detection	Not Hope	4,701	799	773
			Generalized Hope	1,151	186	206
			Unrealistic Hope	546	91	96
			Realistic Hope	505	74	77
	English	2.a) Binary Hope Speech detection	Not Hope	3,104	530	541
			Hope	3,088	502	491
		2. b) Multiclass Hope Speech detection	Not Hope	3,088	502	491
			Generalized Hope	1,726	300	309
			Unrealistic Hope	730	128	124
			Realistic Hope	648	102	108

unrealistic hope, depending on whether hope is based on expectations, desirable facts and undesirable facts respectively.

People from marginalized groups, such as women, LGBTIQ+, racial minorities, physically challenged and so on, who are usually the main targets of abusive and offensive content and other people on social media platforms who are discouraged or in distress, will be very helpful if they get motivated positively through these platforms [4]. While hate speech, fake news, offensive and abusive language, are associated with negativity, hope speech is associated with positivity with an aim of bringing a ray of hope in one's life [5, 2]. Thus detecting hope speech is important for mental health, combating discrimination, and fostering peaceful environments for social media users. Identifying hope speech content in social media amplifies positivity, curbing despair's spread and is vital in balancing online discourse, countering negativity, and fostering resilience. It also empowers users to navigate social media safely combating the harmful content [6]. However, hope speech detection has to be carried out in conjunction with hate speech detection. Otherwise, it might lead to prejudice and individuals who make hurtful and injurious remarks on the internet might act erratically [7].

Manual detection of hope speech on social media is a cumbersome and time consuming process due to the increasing number of social media users as well as hope speech content. This has given rise to the development of tools and techniques which can automatically detect hope speech content. To address the challenges of identifying hope speech in social media platforms, "HOPE at IberLEF 2024" shared task organized at IberLEF 2024, invites the research community to develop models to detect hope speech in Spanish and English languages [8]. The shared task consists of two tasks: Task 1 - Hope

Table 2
Sample texts and their corresponding labels for datasets in Spanish

Task	Sample Text	Translated Text	Label
Task 1: HopeEDI	Sigue siendo necesario luchar x un mundo en el q todxs podamos sentirnos libres. D ser, d amar y d elegir ntros vínculos #OrgulloLGTBI	It is still necessary to fight for a world in which we can all feel free. D be, d love and d choose our links #LGTBIPride	hs
	famosos: AMAD a quien querais la T de LGTB significando Tarradellas para ellos #OrgulloLGTBI #pride #Orgullo #Orgullo2021	celebrities: LOVE whoever you want the T in LGTB meaning Tarradellas for them #OrgulloLGTBI #pride #Orgullo #Orgullo2021	nhs
Task 2: Hope as Expectations			
Subtask: 2.a) Binary Hope Speech Detection	Creo que estamos ante el mejor episodio de esta primera temporada. #TheLastOfUs	I think this is the best episode of this first season. #TheLastOfUs	Not Hope
	Mientras me persigno y le rezo a la Virgen delCarmen. Así e. #URL#	While I cross myself and pray to the Virgin of Carmen. That's how it is. #URL#	Hope
Subtask: 2.b) Multiclass Hope Speech Detection	¿Tu dolor o anhelo me afecta? ¿NO? Pues sufre en silencio, yo no te debo nada. A mamarla. #URL#	Does your pain or longing affect me? NO? Well, suffer in silence, I don't owe you anything. To suck it. #URL#	Not Hope
	No puedo esperar a estar en la playa con mis amigas, tomando el sol, llenándome de arena, comiendo y bebiendo rico.	I can't wait to be at the beach with my friends, sunbathing, getting covered in sand, eating and drinking delicious food.	Unrealistic Hope
	Y #USER# ha cumplido este finde ese sueño por mi.	And #USER# has fulfilled that dream for me this weekend.	Generalized Hope
	q suerte q mí papá me trajo a la escuela y me ahorre esperar el bondi con tremendo sol	I'm lucky that my dad brought me to school and saved me from waiting for the bondi with tremendous sun.	Realistic Hope

for Equality, Diversity and Inclusion (HopeEDI) and Task 2 - Hope as Expectations, consists of two subtasks. Description of these tasks along with the data distribution is shown in Table 1 and sample text from the given dataset for Spanish and English are shown in the Tables 2 and 3 respectively. The code to reproduce the proposed models is available in github[1]. To explore the strategies of detecting hope speech in Spanish and English on social media platforms, in this paper, we - team MUCS, describe the models submitted to the "HOPE at IberLEF 2024" shared task. The details of the proposed models along with the features used to train the models are shown in Table 4.

The rest of paper is organized as follows: Section 2 describes the recent literature on hope speech detection and Section 3 focuses on the description of the proposed models followed by the experiments and results in Section 4. The paper concludes with future works in Section 5.

2. Related Work

Hope speech detection refers to the automatic identification of positive and supportive messages, particularly in online communication platforms like social media, to promote positive and inclusive communication. There have been numerous studies exploring the importance of hope speech detection

[1]https://github.com/SonithD/MUCS-hopeIberLEF2024

11.5

Table 3
Sample texts and their corresponding labels for datasets in English

Task	Sample Text	Label
Subtask: 2.a) Binary Hope Speech Detection	#USER# Oh shit really? I would hope they'd shed some more light on what happened to him in the future	Hope
	#USER# So how many elections has it been now that "everything is on the line"? I've lost count	Not Hope
Subtask: 2.b) Multiclass Hope Speech Detection	I WISH I CHARGED YM FUCKING APPLE WATCH #URL#	Not Hope
	Hope that pool party doesnt involve the guys i am Not strong enough #URL# #URL#	Unrealistic Hope
	Hello #USER# Orlando hoping to hear #YetToCome #USER# on your station today Thank you very much for spreading the "yet to come"	Generalized Hope
	I lean so much more towards the masc side of the gender spectrum(?) and im expected to be this perfect daughter, the other day i was watching a music video with my partner and i was focusing on the guys voice just wishing and hoping that i could somehow have a deep voice	Realistic Hope

across different languages in fostering positive interactions and preventing negativity online. Few of the relevant works are described below:
Balouchzahi et al. [9] developed datasets for binary and multiclass classification, to identify hope content in English tweets and used multiple baselines (ML, Deep Learning (DL), and TL techniques) to benchmark their datasets. TF-IDF of words are used to train ML models (SVM, Decision Tree (DT), Random Forest (RF), LR, XGB, MLP, CatBoost (CB)), Global Vectors for Word Representation (GloVe) and FastText embeddings are used to train DL models (Long Short Term Memory (LSTM), Bidirectional LSTM, and Convolutional Neural Network (CNN), and TL based models are trained using fine-tuned BERT, Robustly Optimised BERT Approach (RoBERTa), and multilingual BERT (mBERT). Among all the learning models, LR and CB classifiers outperformed other classifiers with macro F1 scores of 0.80 and 0.79 for binary classification and 0.64 and 0.54 for multiclass classification respectively. Chakravarthi [7] constructed a hope speech dataset for binary classification covering Equality, Diversity and Inclusion (HopeEDI) containing 28,451, 20,198 and 10,705 user generated comments in English, Tamil and Malayalam languages respectively, from YouTube comments. They experimented with various ML (multinomial Naive Bayes (MNB), k-Nearest Neighbors (kNN), SVM, DT, LR) models trained with TF-IDF of word unigrams. Among their proposed models, DT models exhibited the best macro F1 scores of 0.46 and 0.56 for English and Malayalam languages respectively, and LR model exhibited the best macro F1 score of 0.55 for Tamil language.

Sidorov et al. [10] proposed TL based models using Simpletransformers (BERT, A Lite BERT (ALBERT), RoBERTa, Distilled version of BERT (DistilBERT), XLNet, and Efficiently Learning an Encoder that Classifies Token Replacements Accurately (ELECTRA)) fine tuned with regret and hope speech detection in English with ReDDIT dataset consisting of three labels ('No Regret', 'Regret by Action', and 'Regret by Inaction') and PolyHope dataset with four labels ('Not Hope', 'Generalized Hope', 'Realistic Hope', and 'Not Hope'). Among their proposed models, RoBERTa achieved the highest performance for regret detection, with an averaged macro F1-score of 0.83 and the BERT model for PolyHope dataset outperformed the rest of the models with an averaged macro F1-score of 0.72. Shahiki-Tash et al. [11] proposed 5-layer CNN model trained with keras embeddings to classify Spanish and English texts into Hope or Non-Hope categories and obtained macro F1 scores of 0.4974 and 0.7238 for Spanish and English texts respectively. Balouchzahi et al. [12] presents an ensemble model with soft voting to select the best word and character n-grams to train keras Neural Network (NN) for hope speech detection

Table 4
Learning models and the features used to train the models

Languages	Subtasks	Models	Features
Task 1: Hope for Equality, Diversity and Inclusion			
Spanish	Hope Speech detection (Train set consists of text in LGTBI domain and test set in unknown domains)	ML_SVM	TF-IDF of word n-grams in the range (1, 3), Multilingual word embedding, Aligned word vectors
		ML_LR	
		ML_LSVC	
		ML_RF	
		ML_CatBoost	
		ML_XGBoost	
		ML_AdaBoost	
		TL_DistilSpanbert	DistilSpanBERT
		TL_mBERT	mBERT
		Hope_probfuse	Sentence Transformers
Task 2: Hope as Expectations			
Spanish	2.a) Binary Hope Speech detection	TL_DistilSpanBERT	DistilSpanBERT
		TL_SpanBERT	SpanBERT
		Hope_probefuse	Sentence Transformers
	2. b) Multiclass Hope Speech detection	ML_SVM	Aligned word vectors
		ML_LR	
		TL_DistilSpanBERT	DistilSpanBERT
English	2.a) Binary Hope Speech detection	ML_SVM	TF-IDF of word n-grams in the range (1, 3)
		ML_RF	
		ML_LSVC	
		ML_LR	
		ML_XGBoost	
		ML_AdaBoost	
		logloss_RF	
		TL_BERT	BERT
	2. b) Multiclass Hope Speech detection	ML_LSVC	TF-IDF of word n-grams in the range (1, 3)
		Hope_probfuse	Sentence Transformers
Multilingual models for Spanish and English	2. b) Multiclass Hope Speech detection	MLM_LR (MUSE Emb+LR)	Multilingual word embedding
		MLM_LR (alignvec+LR)	Aligned word vectors
		TL_Ensemble	Variants of BERT

and obtained weighted F1 scores of 0.790 and 0.870 for Spanish and English texts respectively.

Puranik et al. [13] proposed two models: i) CNN model with dense layer trained with several BERT variants (bert-base-uncased, albert-base, distilbert-base-uncased, roberta-base, character-bert, Universal Language Model Fine-tuning (ULMFiT) for English text, mbert-uncased, mbert-cased, indic-bert, xlm-roberta-base, distilmbert-cased, MuRIL for Malayalam and Tamil code-mixed texts), and ii) Bidirectional Long Short-Term Memory (BiLSTM) trained with several BERT variants (bert-base-cased for English text and mbert-uncased, mbert-cased, xlm-roberta-base, Multilingual Representations for Indian Languages (MuRIL) for Malayalam and Tamil texts respectively), for hope speech detection. Their proposed CNN dense model trained with ULMFiT obtained weighted F1-score of 0.94 for English and BiLSTM model trained with mbert-uncased obtained 0.8545 for Malayalam and dense model trained with distilmBERT-cased obtained 0.59 weighted F1-score for Tamil code-mixed text. Aggarwal et al. [14] experimented with ML models (Naive Bayes, LR and SVM) and BERT model on relabelled data for hope speech detection on social media platforms and their BERT model exhibited the best macro F1 score of 0.85.

The literature highlights extensive research efforts aimed at detecting hope speech across languages like English, Spanish, Tamil, and Malayalam. These studies utilize a range of ML, DL, and TL models, offering valuable insights into their work. However, not all results are promising and this encourages us to explore more models for hope speech detection.

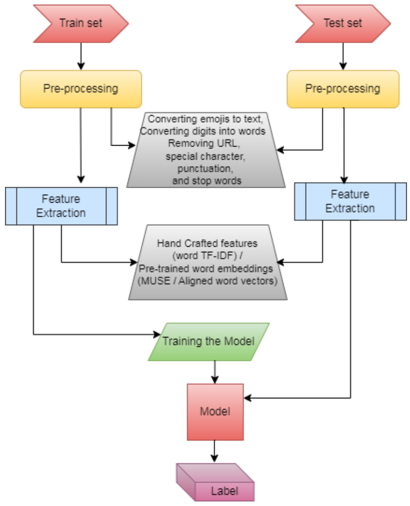


Figure 1: Framework of the proposed ML model

3. Methodology

Data imbalance occurs when there is a large variation in the number of instances within the given classes. The datasets provided by the shared task organizers are imbalanced and this influences the learning models to be biased towards the majority class while performing poor for the minority class. This bias could be reduced to some extent by balancing the dataset either by increasing the minority class samples or decreasing the majority class samples. Two distinct techniques used in this work to balance the datasets are described below:

- **Random oversampling**[2] - is an oversampling technique which increases the instances in the minority class by replicating the samples in the minority class without adding any new data. This technique increases the sample size to match the number of samples in the majority class, thus balancing the data.
- **Natural Language Processing Augmentation (NLPAug)**[3]- entails creating new samples from the existing data by transforming or augmenting it in different ways such as - synonym replacement, word insertion, and word deletion. Augmentation will increase the diversity of datasets and robustness of learning models. *TfIdfAug* in NLPAug with options - "insert", "substitute", and "SynonymAug" sourced from WordNet, are applied to the minority class to increase the sample size to match the number of samples in the majority class.

The proposed methodology includes balancing the imbalance dataset followed by constructing the learning models. We have explored ML models and TL techniques for hope speech detection in Spanish and English and the steps involved in the construction of these models are explained in the following sections.

3.1. Machine Learning Models

The framework of ML models is visualized in Figure 1 and the steps included in the construction are explained below:

[2]https://imbalanced-learn.org/stable/over_sampling.html
[3]https://pypi.org/project/nlpaug/0.0.5/

3.1.1. Pre-processing

Pre-processing encompasses various techniques to remove noise from the text and normalize the text, with the aim of improving the performance of the learning models. As emojis depict user's thoughts, they are converted to text using demoji[4] library and numeric information is converted to words. URLs, user mentions, hash tags, special characters, and punctuation, present in the text do not contribute to the classification task and hence are removed. Stopwords are a set of commonly used words in a language and they do not contribute significantly to the classification task and hence are removed. As the dataset provided for the shared task includes Spanish and English words, Spanish and English stopwords available at the Natural Language Tool Kit (NLTK)[5] library are used as references to remove Spanish and English stopwords respectively.

3.1.2. Feature Extraction

The role of feature extraction is to extract distinguishable features from the given text with the objective of improving the performance of the learning models. The following features are extracted from the given text:

- **Hand-crafted features**: Word n-grams is a sequence of n contiguous words in a given text and is a language independent feature and word ngrams in the range (1, 3) are obtained from the input text and converted to TF-IDF vectors using TfidfVectorizer[6]. TF-IDF vectors provide normalised representation of text documents by mitigating the influence of excessively repeated words. These vectors indicate the significance of a word within a specific document relative to the entire corpus.
- **Pretrained Word Embeddings**: Word embeddings is a numerical representation of words in a continuous vector space in which words with similar meanings are grouped together. The importance of word embeddings lies in their ability to capture semantic similarities and the relationships between words based on their usage in large corpora of text. In this work, Multilingual Word Embeddings and Aligned word vectors are used to represent the text and their description is given below:
  - **Multilingual Word Embeddings** - Word embeddings are monolingual by default. Offlate bilingual and multilingual word embeddings are also being developed to support bilingual and multilingual NLP respectively. Multilingual word embeddings offer a valuable resource for building multilingual models by aligning word embeddings from different languages into a shared vector space. These embeddings capture semantic similarities across languages, facilitating effective knowledge transfer and enhancing the performance of multilingual NLP tasks [15, 16]. Multilingual Unsupervised and Supervised Embeddings[7] (MUSE) repository from Facebook provides a comprehensive toolkit for training and evaluating multilingual word embeddings enabling cross-lingual analysis and this is used to represent the given Spanish and English text.
  - **Aligned word vectors**[8] - are derived from word embeddings of multiple languages that have been aligned into a shared vector space. This alignment process ensures that similar concepts across different languages are represented by nearby vectors in the shared space, facilitating effective knowledge transfer between languages [17, 18]. In this study, aligned vector files for Spanish and English languages are leveraged to represent the given text. Combining the aligned vector files of Spanish and English into a unified dictionary creates a comprehensive resource that captures semantic similarities across these two languages and this combined dictionary serves as the foundation for training the multilingual models.

[4]https://pypi.org/project/demoji/
[5]https://pythonspot.com/nltk-stop-words/
[6]https://scikit-learn.org/stable/modules/generated/sklearn.feature_extraction.text.TfidfVectorizer.html
[7]https://github.com/facebookresearch/MUSE
[8]https://fasttext.cc/docs/en/aligned-vectors.html

These sophisticated linguistic representations enhances the model's ability to capture semantic similarities and relationships between the languages.

- **Sentence Transformers**[9] - is a framework that offers a straightforward dense vector representation for the given sentences/text. The framework offers a large collection of pre-trained models like BERT, RoBERTa, and XLM-RoBERTa and achieve state-of-the-art performance across various tasks.

These features are used to train the ML models individually.

3.1.3. Model Construction

The following ML models are used in this work to detect hope speech in Spanish and English:

- **ML_LR** - LR model strategically incorporates dependent variables and regularization techniques to safeguard against over-fitting. It aggregates the features through linear combination followed by the transformation using the logistic function - a process that empowers the algorithm to generate predictions and classify instances into one of the predefined classes [19].
- **ML_SVM** - SVM is highly popular for its effectiveness in high-dimensional feature space making it particularly well-suited for text classification tasks. It's ability to identify intricate and nonlinear relationships between the features allows it to excel in accurately categorizing text documents [20].
- **ML_LSVC** is a LSVC in Scikit-learn library[10] which attempts to maximize the distance between classified samples by finding a hyperplane.
- **ML_RF** - RF is one of the supervised learning algorithms that is flexible, can be adapted easily to different situations, and can be used without any hyperparameters. It is necessary to build a minimum number of trees in order to classify the data [21].
- **ML_kNN** - kNN algorithm ranks the given unlabeled sample's nearest neighbors among the training documents, and use the class labels of k most similar neighbors to predict the class of the given unlabeled samples [22]. The number of neighbors is set to 3 in this study.
- **ML_CB** - CB is a powerful ML algorithm that iteratively builds an ensemble of decision trees, each focusing on different aspects of the data, to collectively make accurate predictions across multiple categories. In this study, binary cross-entropy loss function which effectively penalizes the model for misclassification is used. This loss function also guides towards minimizing the discrepancy between predicted and actual class probabilities [23].
- **ML_XGBoost** - XGBoost is a highly effective ensemble learning algorithm used for classification and regression tasks. It sequentially builds a series of decision trees, refining predictions by focusing on the errors of previous trees. The ensemble boosting technique is employed to optimize the model, using a Taylor expansion to approximate the loss function [24].
- **ML_AdaBoost** - AdaBoost classifier sequentially trains a series of weak classifiers focusing on instances misclassified by the previous classifiers thus enhancing the overall model accuracy. Leveraging AdaBoost as a single base classifier allows to harness its boosting capabilities to effectively improve predictive performance [25].
- **Logloss_RF** - also known as Binary Cross-Entropy Loss is an ensemble learning method that constructs multiple decision trees during training and outputs the mode of the classes for classification tasks or the mean prediction for regression tasks. By computing log loss, the RF classifier's performance can be evaluated, where lower values indicate better alignment between predicted probabilities and true labels, ensuring more accurate classification results.
- **Hope_probfuse** - is an ensemble model with soft voting. In ensembling, for every base model soft voting assigns each class a probability score and the final prediction is then determined by considering the maximum probability of all the base models. In this work, SVM and RF classifiers

[9]https://sbert.net/, https://pypi.org/project/sentence-transformers/
[10]https://scikit-learn.org/stable/modules/generated/sklearn.svm.LinearSVC.html

Table 5
Hyperparameters and their values used in ML models

Model	Hyperparameters	Values
ML_LSVC	C	1.0
	class_weight	balanced
	max_iter	10000
	random_state	123
ML_SVM	kernel	rbf
	random_state	42
ML_LR/ MLM_LR	max_iter	1000
ML_RF	n_estimators	100
	random_state	42
ML_XGBoost	objective	binary:logistic
	random_state	42
ML_AdaBoost	n_estimators	50
	random_state	42
ML_CatBoost	iterations	1000
	learning_rate	0.1
	loss_function	Logloss
	verbose	100

are trained using the vectors represented by Sentence Transformers (ST): STSSPAN[11] and BERTIN [26] respectively, for Spanish datasets and for English, all-distilroberta-v1[12] and IndicSBERT-STS[13] [27] are used to train SVM and RF classifiers respectively.

- **Multilingual Models** - allow to handle multiple language datasets simultaneously. To improve the performance and generalization, these models utilize the shared representations across languages. **MLM_LR** is trained with multilingual word embeddings and aligned word vectors independently to detect hope speech in both Spanish and English. This approach capitalized on the varied linguistic representations captured by multilingual embeddings and aligned vectors, facilitate accurate language-specific predictions while accommodating the multilingual characteristics of the input data.

The hyperparameters and their values used in ML models are shown in Table 5.

3.2. Transfer Learning

In TL, the knowledge obtained from learning the source task is transferred to the target task to speed up learning and improve the performance of the target task rather than starting the target task from scratch [28]. The framework of the proposed TL based model is shown in Figure 2. In this technique, the raw text is pre-processed and transformed into a consistent format by converting the emojis to corresponding text using demoji library, converting numeric information to corresponding words, and removing URLs, user mentions, hash tags, and special characters. These are applied to the sentences of the given text retaining the sentence structure in the text.

Pretrained models are trained on large unlabeled datasets and several pretrained models are available to use for various applications. TL techniques fine-tune the pretrained models on the datasets of the required tasks so that the rich linguistic representations acquired during pretraining are leveraged to improve the performance of the models. BERT[14] is a pretrained model trained on Toronto Book Corpus and Wikipedia and exclusively used for tasks involving English texts whereas mBERT[15] is trained on

[11]https://huggingface.co/hiiamsid/sentence_similarity_spanish_es
[12]https://huggingface.co/sentence-transformers/all-distilroberta-v1
[13]https://huggingface.co/l3cube-pune/indic-sentence-similarity-sbert
[14]https://huggingface.co/google-bert/bert-base-uncased
[15]https://huggingface.co/google-bert/bert-base-multilingual-cased

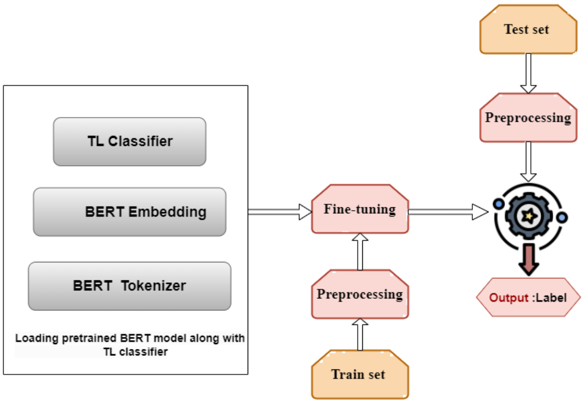


Figure 2: Framework of the proposed TL based model

Table 6
Performances of the proposed ML and TL based models in Task 1: HopeEDI - Binary Hope Speech detection in Spanish

Model	Val Set			Test Set		
	Precision	Recall	F1 score	Precision	Recall	F1 score
ML models						
ML_SVM	0.76	0.76	0.75	0.60	0.58	0.55
ML_LSVC	0.74	0.72	0.71	0.60	0.59	**0.59**
ML_LR	0.75	0.74	0.74	0.60	0.58	0.56
ML_RF	0.73	0.69	0.68	0.63	0.57	0.50
ML_kNN	0.66	0.65	0.65	0.53	0.53	0.53
ML_CatBoost	0.71	0.71	0.71	0.63	0.57	0.52
ML_XGBoost	0.69	0.69	0.68	0.61	0.57	0.54
ML_AdaBoost	0.62	0.61	0.61	0.64	0.59	0.55
Hope_probfuse	0.83	0.80	0.80	0.62	0.59	0.56
TL based models						
TL_DistilSpanbert	0.84	0.83	0.83	0.60	0.59	0.58
TL_mBERT	0.73	0.65	0.61	0.56	0.52	0.43
F1 score: Macro F1 score						

wikipedia data and blogs that belong to more than 104 languages including English and Spanish and exclusively used for tasks that include multiple languages [29]. Similarly, SpanishBERT[16] is trained on the Spanish edition of Wikipedia, the OPUS Project, and Spanish books and news articles, and is exclusively used for tasks involving Spanish texts. Further, DistilSpanBERT[17] is a distilled version of SpanishBERT, trained on Spanish text sources, and is also exclusively used for tasks involving Spanish texts, but is optimized for efficiency and speed. In this study, **TL_SpanBERT**, **TL_DistilSpanBERT**, and **TL_BERT** fine-tunes SpanishBERT, DistilSpanBERT, and BERT base model, respectively.

3.2.1. Multilingual Models

The pre-processed Spanish and English datasets are combined to form a unified dataset to effectively process linguistic information from both the languages simultaneously. The combined dataset is then used to fine-tune the pretrained multilingual models and build the transformer classifier (Classification-Model) to make the predictions for Spanish and English text. The following multilingual models are explored for handling Spanish and English text together:

[16]https://huggingface.co/dccuchile/bert-base-spanish-wwm-cased
[17]https://huggingface.co/dccuchile/distilbert-base-spanish-uncased

Table 7
Performances of the proposed TL based models in subtask 2.a) - Binary Hope Speech detection in Spanish

Model	Val Set			Test Set		
	Precision	Recall	F1 score	Precision	Recall	F1 score
Hope_probefuse	0.77	0.76	0.76	0.78	0.77	0.77
TL_DistilSpanBERT	0.80	0.82	0.81	0.81	0.82	0.82
TL_SpanBERT	0.80	0.80	0.80	0.82	0.82	**0.82**
F1 score: Macro F1 score						

Table 8
Performances of the proposed ML and TL based models in subtask 2.a) - Binary Hope Speech detection in English

Model	Val Set			Test Set		
	Precision	Recall	Macro F1 score	Precision	Recall	Macro F1 score
ML models						
ML_SVM	0.79	0.79	0.79	0.81	0.81	0.81
ML_LSVC	0.81	0.81	0.81	0.81	0.81	0.81
ML_LR	0.79	0.79	0.79	0.81	0.81	0.81
ML_RF	0.77	0.77	0.77	0.78	0.76	0.76
ML_kNN	0.64	0.63	0.63	0.64	0.64	0.64
ML_XGBoost	0.79	0.79	0.79	0.80	0.79	0.79
ML_CatBoost	0.81	0.81	0.81	0.82	0.82	0.82
ML_AdaBoost	0.79	0.79	0.79	0.78	0.78	0.78
logloss_RF	0.79	0.79	0.79	0.80	0.79	0.80
TL based models						
TL_BERT	0.81	0.81	0.81	0.82	0.82	**0.82**

Table 9
Performances of ML and TL based models in subtask 2.b) - Multiclass Hope Speech detection in Spanish

Model	Val Set						Test Set		
	P	R	Macro F1 score	P	R	Macro F1 score	P	R	Macro F1 score
ML models	**Without oversampling**			**With oversampling**					
ML_SVM	0.43	0.49	0.44	0.44	0.50	0.46	0.26	0.26	0.25
ML_LR	0.48	0.32	0.32	0.43	0.51	0.45	0.42	0.43	0.41
TL based model	**Without augmentation**			**With augmentation**					
TL_Distil SpanBERT	0.54	0.48	0.49	0.56	0.50	0.51	0.63	0.65	**0.64**
P: Precision, **R**: Recall									

- **TL_mBERT** - is very effective for multilingual NLP tasks because of its thorough pretraining on more than 104 languages, which enables it to capture cross-linguistic patterns and semantic similarities from multiple languages simultaneously. mBERT model is fine-tuned by the combined dataset.
- **TL_Ensemble** - is an ensemble model comprising of three pretrained transformer models: google-bert/bert-base-multilingual-cased, bert-base-uncased, and distilbert/distilbert-base-uncased[18], with the aim of improving the performance of the classifier. Each model is fine-tuned by the combined dataset and is used to train transformer classifier (ClassificationModel) to make the predictions for Spanish and English text. Further, the predictions of the ensemble models are combined by averaging the raw output logits followed by applying softmax to derive the final

[18]https://huggingface.co/distilbert/distilbert-base-uncased

Table 10
Peformances of the proposed models in subask 2.b) - Multiclass Hope Speech detection in English

Model	Val set			Test set		
	Precision	Recall	Macro F1 score	Precision	Recall	Macro F1 score
ML_LSVC	0.57	0.57	0.57	0.56	0.58	0.57
Hope_probfuse	0.53	0.54	0.53	0.29	0.38	0.32

Table 11
Performances of Multilingual models in subtask 2.b) - Multiclass Hope Speech detection in Spanish and English for Validation set

Model	Precision	Recall	Macro F1 score	Precision	Recall	Macro F1 score
	Without oversampling			With oversampling		
MLM_LR (MUSE Emb+LR)	0.70	0.38	0.38	0.50	0.56	0.51
MLM_LR (alignvec+LR)	0.75	0.35	0.34	0.47	0.52	0.47
TL_mBERT	0.57	0.50	0.52	0.57	0.50	0.52
TL_Ensemble	0.56	0.46	0.47	-	-	-

Table 12
Performances of Multilingual models in subtask 2.b) - Multiclass Hope Speech detection in Spanish and English for Test set

Model	English			Spanish		
	Precision	Recall	Macro F1 score	Precision	Recall	Macro F1 score
MLM_LR (MUSE Emb+LR)	0.50	0.55	0.50	0.49	0.52	0.49
MLM_LR (alignvec+LR)	0.25	0.24	0.23	0.42	0.43	0.41
TL_mBERT	0.62	0.60	0.61	0.59	0.65	0.61
TL_Ensemble	0.59	0.54	**0.56**	0.49	0.39	0.38

Table 13
Results of our best performed models with macro F1 score and Ranks for Spanish and English datasets

Task	Model	F1 score	Rank
Task 1: HopeEDI	ML_LSVC	0.59	10
Subtask 2.a) Binary Hope Speech detection in Spanish	TL_SpanBERT	0.82	5
Subtask 2.a) Binary Hope Speech detection in English	TL_BERT	0.82	9
Subtask 2.b) Multiclass Hope Speech detection in Spanish	TL_DistilSpanBERT	0.64	4
Subtask 2.b) Multiclass Hope Speech detection in English	TL_Ensemble	0.56	8
F1 score: Macro F1 score			

class probabilities.

Instead of having individual models for each language, having a single multilingual model for multiple languages will be more economic and time saving.

4. Experiments and Results

Various experiments were carried out on the datasets provided by the shared task organizers, using different techniques for balancing the dataset, different combinations of features, and different learning models, to identify the hope speech in Spanish and English. The models which gave good results on the Validation (Val) set are used to predict the labels of the Test set. Random oversampling is applied for: i) subtask 2.a) Binary Hope Speech detection in Spanish language to train Hope_probfuse model and ii) subtask 2.b) Multiclass Hope Speech detection, both in Spanish and English, to train

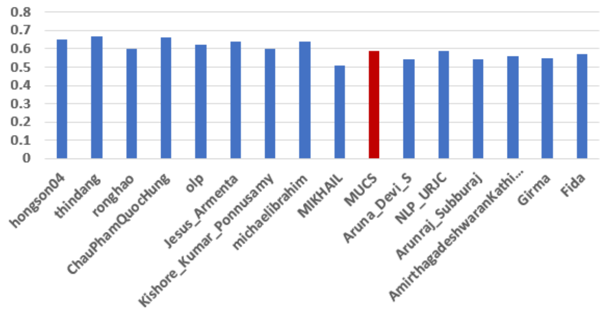
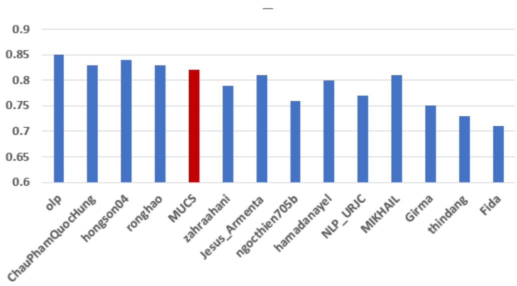
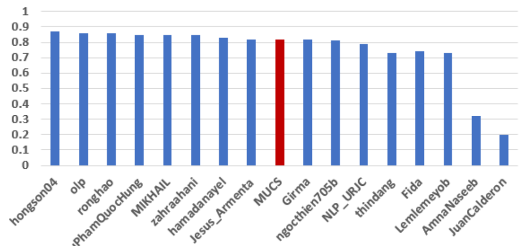


Figure 3: Comparison of macro F1 scores of the participating teams in Task 1 HopeEDI - Binary Hope Speech detection in Spanish

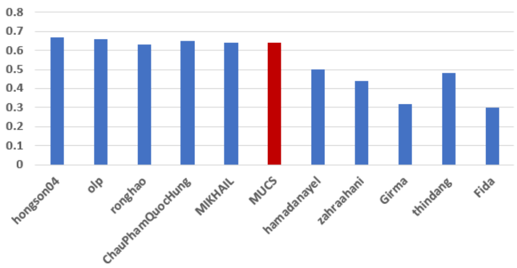


(a) *Spanish*

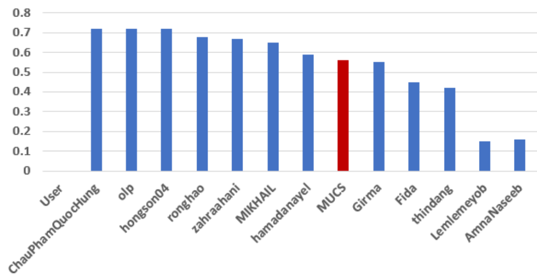


(b) *English*

Figure 4: Comparison of macro F1 scores of the participating teams in subtask 2a - Binary Hope Speech detection in Spanish and English



(a) *Spanish*



(b) *English*

Figure 5: Comparison of macro F1 scores of the participating teams in subtask 2a - Mulitclass Hope Speech detection in Spanish and English

Hope_probfuse, MLM_LR, and TL_mBERT models. The NLPAug is applied for subtask 2. b) Multiclass Hope Speech detection task both in Spanish and English, to train TL_DistilSpanBERT and ML_LSVC models, respectively.

The performances of the models are evaluated by the organizers based on macro F1 scores. The performances of the proposed models on the Val and Test set for Task 1: HopeEDI is shown in Table 6. The performances of proposed models for subtask 2.a) Binary Hope Speech detection in Spanish and English and subtask 2. b) Multiclass Hope Speech detection in Spanish and English, are shown in Table 7, Table 8, Table 9, and Table 10 respectively. Further, the performances of multilingual models in subtask 2.b) - Multiclass Hope Speech detection for Validation and Test set are shown in Table 11 and Table 12 respectively. Among our proposed models, the macro F1 scores of the best performing models are shown in Table 13. Figures 3, 4, and 5 gives the comparison of macro F1 scores of all the participating teams for the shared Task 1: HopeEDI, subtask, 2.a) Binary Hope Speech detection in Spanish and English, and subtask 2. b) Multiclass Hope Speech detection in Spanish and English respectively.

Few samples of misclassified comments along with the actual and predicted labels obtained from

Table 14
Samples of misclassification in subtask 2.a) - Binary Hope Speech detection in English

Text	Actual Label	Predicted Label	Remark
#USER# Yeah people got mad at me for posting these too much on main.	Not Hope	Hope	The term "posting" is generally connected with activities expressing involvement or participation, which may convey hope or positivity and hence it may be categorirzed as hope.
It's just so hard to believe. I'll be ending my term by Thursday just like VP Leni. We may not have won the battle but we will continue to reign and win the war.	Hope	Not Hope	The model might have focused on the term "hard to believe", which appears negative and hence have missed the overall hopeful message in the text.

TL_BERT for subtask 2.a) Binary Hope Speech detection in English and TL_Ensemble models for subtask 2.b) Multiclass Hope Speech detection in English, and the probable reasons for misclassification are shown in Table 14 and 15 respectively.

5. Conclusion and Future Work

In this paper, we - team MUCS, describe the models submitted to Hope Speech detection shared task at IberLEF 2024 to identify hope speech in Spanish and English. The experiments are conducted with two techniques to handle data imbalance - random oversampling and NLPAug, and handcrafted features (TF-IDF of words n-grams in the range (1, 3)), multilingual word embeddings, and aligned word vectors, are used to train the ML classifiers individually. Multilingual ML models are trained using multilingual embeddings and aligned word vectors and BERT variants are fine-tuned in multilingual TL based models. Among the proposed models, ML_LSVC, TL_SpanBERT, TL_BERT, TL_DistilSpanBERT , and TL_Ensemble models obtained 10th, 5th, 9 th, 4th, and 8th ranks, by exhibiting macro F1 scores of 0.59, 0.82, 0.82, 0.64, and 0.56 for Task 1: HopeEDI in Spanish, subtask 2.a) Binary Hope Speech detection in Spanish and English, and subtask 2.b) Multiclass Hope Speech detection in Spanish and English respectively. Efficient feature combinations and different learning approaches will be explored further.

References

[1] S. M. Kavatagi, R. R. Rachh, S. S. Biradar, VTUBGM@ LT-EDI-2023: Hope Speech Identification using Layered Differential Training of ULMFit, in: Proceedings of the Third Workshop on Language Technology for Equality, Diversity and Inclusion, 2023, pp. 209–213.
[2] M. G. Yigezu, G. Y. Bade, O. Kolesnikova, G. Sidorov, A. Gelbukh, Multilingual Hope Speech Detection using Machine Learning (2023).
[3] 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).
[4] 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).
[5] 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.
[6] D. García-Baena, M. Á. García-Cumbreras, S. M. Jiménez-Zafra, J. A. García-Díaz, R. Valencia-

Table 15
Samples of misclassification in subtask 2.b) - Multiclass Hope Speech detection in English

Text	Actual label	Predicted label	Remark
#USER# Hehehe We will definitely do that.. I think I will be homeless soon I hope you will be available to accommodate a homeless guy	Realistic Hope	Generalized Hope	This may be due to the model's inability to accurately contextualize the word 'homelessness' within the broader optimistic sentiment expressed in the sentence, leading to an incorrect categorization.
#USER# #USER# #USER# Where are you getting this info from? There is nothing that can be done here other than hoping individuals can grow and be I can't change anything that's been said, but I can voice my opinion on the community as a whole when it's been attacked.	Generalized Hope	Not Hope	This could be due to the model's inability to distinguish nuanced expressions of hope from the text containing negative or resigned phrases: "nothing that can be done" and "can't change anything". These might have biased the model towards interpreting the text as lacking hope, leading to the incorrect classification.
Was hoping to finish this tomorrow, but I got emergency-called to work for a couple hours. Will see ;_; also this tropical heat is killing me #URL#	Not Hope	Realistic Hope	This could be attributed to the model's failure to recognize the expression of frustration and disappointment conveyed by the mention of 'emergency-called to work' and discomfort from 'tropical heat' leading to inaccurate categorization.
#USER# damn man i'm here just tryna be a nice friend and dis is how u treat me i hope u choke with cold spaghetti	Unrealistic Hope	Generalized Hope	This might be due to the model's inability to adequately capture the negative sentiment expressed in the text. The model may have overlooked the hostile tone and interpreted the mention of "tryna to be a nice friend" as an expression of Generalized Hope, leading to an incorrect categorization.

García, Hope Speech Detection in Spanish: The LGBT case, Language Resources and Evaluation 57 (2023) 1487–1514.
[7] B. R. Chakravarthi, Hope Speech Detection in YouTube Comments, Social Network Analysis and Mining 12 (2022) 75.
[8] 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.
[9] F. Balouchzahi, G. Sidorov, A. Gelbukh, PolyHope: Two-level Hope Speech Detection from Tweets, in: Expert Systems with Applications, Elsevier, 2023, p. 120078.
[10] 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.
[11] M. Shahiki-Tash, J. Armenta-Segura, O. Kolesnikova, G. Sidorov, A. Gelbukh, Lidoma at hope2023iberlef: Hope Speech Detection using Lexical Features and Convolutional Neural Networks, in: Proceedings of the Iberian Languages Evaluation Forum (IberLEF 2023), co-located

with the 39th Conference of the Spanish Society for Natural Language Processing (SEPLN 2023), CEUR-WS. org, 2023.
[12] 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, Association for Computational Linguistics, Dublin, Ireland, 2022, pp. 206–211. doi:10.18653/v1/2022.ltedi-1.28.
[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] P. Aggarwal, P. Chandana, J. Nemade, S. Sharma, S. Saumya, S. Biradar, Hope Speech Detection on Social Media Platforms, 2022. arXiv:2212.07424.
[15] A. Conneau, G. Lample, M. Ranzato, L. Denoyer, H. Jégou, Word Translation Without Parallel Data, arXiv preprint arXiv:1710.04087 (2017).
[16] G. Lample, A. Conneau, L. Denoyer, M. Ranzato, Unsupervised Machine Translation Using Monolingual Corpora Only, arXiv preprint arXiv:1711.00043 (2017).
[17] P. Bojanowski, E. Grave, A. Joulin, T. Mikolov, Enriching Word Vectors with Subword Information, Transactions of the Association for Computational Linguistics 5 (2017) 135–146.
[18] A. Joulin, P. Bojanowski, T. Mikolov, H. Jégou, E. Grave, Loss in Translation: Learning Bilingual Word Mapping with a Retrieval Criterion, in: Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, 2018.
[19] S. U. Hassan, J. Ahamed, K. Ahmad, Analytics of Machine Learning-Based Algorithms for Text Classification, in: Sustainable Operations and Computers, Elsevier, 2022, pp. 238–248.
[20] N. Kalcheva, M. Karova, I. Penev, Comparison of the Accuracy of SVM Kemel Functions in Text Classification, in: 2020 International Conference on Biomedical Innovations and Applications (BIA), IEEE, 2020, pp. 141–145.
[21] M. Huljanah, Z. Rustam, S. Utama, T. Siswantining, Feature Selection using Random Forest Classifier for Predicting Prostate Cancer, in: IOP Conference Series: Materials Science and Engineering, volume 546, IOP Publishing, 2019, p. 052031.
[22] S. Zhang, X. Li, M. Zong, X. Zhu, D. Cheng, Learning k for knn Classification, ACM Transactions on Intelligent Systems and Technology (TIST) 8 (2017) 1–19.
[23] A. A. Badr, A. K. Abdul-Hassan, CatBoost Machine Learning Based Feature Selection for Age and Gender Recognition in Short Speech Utterances., International Journal of Intelligent Engineering & Systems 14 (2021).
[24] U. Salamah, A Comparison of Text Classification Techniques Applied to Indonesian Text Dataset, Int. J. Sci. Res. Comput. Sci. Eng. Inf. Technol 5 (2019) 217–222.
[25] M. Li, P. Xiao, J. Zhang, Text Classification Based on Ensemble Extreme Learning Machine, arXiv preprint arXiv:1805.06525 (2018).
[26] 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.
[27] S. Deode, J. Gadre, A. Kajale, A. Joshi, R. Joshi, L3Cube-IndicSBERT: A Simple Approach for Learning Cross-Lingual Sentence Representations using Multilingual BERT, arXiv preprint arXiv:2304.11434 (2023).
[28] B. Fazlourrahman, B. Aparna, H. Shashirekha, CoFFiTT-COVID-19 Fake News Detection Using Fine-Tuned Transfer Learning Approaches, in: Congress on Intelligent Systems: Proceedings of CIS 2021, Volume 2, Springer, 2022, pp. 879–890.
[29] A. Hegde, G. Kavya, S. Coelho, H. L. Shashirekha, MUCS@ LT-EDI2023: Learning Approaches for Hope Speech Detection in Social Media Text, in: Proceedings of the Third Workshop on Language Technology for Equality, Diversity and Inclusion, 2023, pp. 279–286.