Comparative Study of Various Techniques for Rude and Threat dialect Detection in Marathi

Bhushan Nikam
School of Computer Sciences
North Maharashtra University,
Jalgaon (MS), India

Nita Patil, PhD
School of Computer Sciences
North Maharashtra University,
Jalgaon (MS), India

ABSTRACT
Rude and threatening language recognitions aim to protect individuals and online communities from harmful and offensive content. It can be applied in various contexts, like comment sections or other online communication social channels. This paper compares various tools and techniques for Abusive and Threat Language Detection in Marathi. The research observations of the methods, strategies, and features needed to implement Marathi abusive and threat language detection are reported.

General Terms
NOT (Non Hate-Offensive), HOF (Hate and Offensive), CNN (Convolution Neural Network), LSTM (Long short-term memory), TF-IDF (Term Frequency - Inverse Document Frequency), FIRE- (Forum for Information Retrieval Evaluation)

Keywords
Transformer, Monolingual, multilingual, algorithms.

1. INTRODUCTION
The importance of studying and observing the detection of abusive and threatening language in the Marathi language lies in protecting individuals, promoting online safety, enhancing content moderation, preserving freedom of speech, and addressing language-specific challenges. This task often involves machine learning algorithms, like supervised, to train models on labelled data. The models gain arrangements and appearance occurrences in the text to predict the presence of abusive or threatening language. Additionally, the task may require techniques from subfields of NLP, such as sentiment analysis, semantic analysis, or language modelling, to capture the nuances and context- specific aspects of abusive and threatening language in the Marathi language. These efforts create a secure and broader digital circumstance for the Marathi language. Section 2 describes the origin of the problems, the review of literature is mentioned in Section 3, the investigation and findings concerning the methods, tools, strategies, and attributes necessary for identifying abusive and threatening language in Marathi are reported in Sections 4 and 5.

2. RELATED WORK
Some efforts have been started to detect offensive and hateful language from social media on other than English Language. Most of them followed Transformer based techniques whereas others used simple machine and deep learning algorithms. Little work is initiated recently to detect threats and abusive Marathi Language on social media, which is one of the most used Language in India.

Gamal et al. [11], reviewed in detail various Cyber-Hate categories, Cyber-Hate speech detection approaches and challenges. With an intension of implementing optimized real-time Arabic cyberbullying detection system, they comparatively studied binary and multiple classifications over different datasets.

Ranasinghe et al. [21] overviewed the Offensive Language Identification task in Marathi at FIRE 2022, observing that traditional approaches are practical for cramped training data. In addition, ensembled framework and cross-lexemic substitutional learning would still feasibly beat traditional techniques' results if training data provided additional extra occurrences in Marathi or Hindi.

In their comparative study of single and multiple-language BERT models for Marathi and Hindi dataset, Velankar et al. [26], stated that models trained in advance did not add general differential sentence vector representations for the classification task. Also, the mono-lingual model offers superior sentence entrenching to the multi-language one, thus emphasizing the requirement for more accurate sentence embedding models.

3. ORIGIN OF THE THREAT AND ABUSIVE MARATHI LANGUAGE DETECTION PROBLEM
To address the new challenges in multilingual information access and retrieval, mainly for the languages of the Indian Subcontinent, the Information Retrieval Research Group of the Indian Statistical Institute, Kolkata, conceptualized and launched the Forum of Information Retrieval (FIRE) in 2008. While significant efforts have already been made to develop hurtful speech recognition models for English, the deficiency of resources and designated data for other languages has challenged researchers and practitioners.

To fix this gap, researchers at the University of Hamburg, Germany, conducted the HASOC project - "Offensive and Hate Speech Content Detection in Indo-European Speeches [18]: Benchmark Data and Classification," which was conducted in 2018 under the umbrella of FIRE. The goal was to create a benchmark dataset for detecting hurtful speech context recognition in several Indo-European languages.

The initiative organized a collective task competition, inviting researchers and practitioners to evolve and assess their harmful speech detection systems using the provided dataset.
The dataset included disgusting speech, harmful language, and non-filthy content in multiple languages, such as English, German, Hindi, and others.

Through the collaborative efforts of researchers and participants, the HASOC drive aimed to improve the understanding of harmful speech and undesirable content in diverse languages and foster the development of effective detection and moderation systems. The shared task

competitions provided a platform for evaluating and comparing different approaches, techniques, and models for harmful speech detection, contributing to advancements in the field.

The HASOC ambition has played an essential role in enhancing consciousness about harmful language and adverse motives in various Indo-European dialects, encouraging research and development in this vital area, and promoting safer online environments for users across different linguistic communities.

In 2019, the first attempt at a Multi-language Cyberbullying Detection System to recognize online harassment behavior in twin Indian dialects – Hindi and Marathi [16], was seen. In recent years, work has progressed by many researchers. HASOC work started on Marathi Language in 2021.

Table 1 Threat and Abusive Marathi Language Identification Evaluation Forums

Conference	Year(s)	Sponsor
FIRE '21	2021	ACM
FIRE '22	2022	ACM

4. LITRATURE REVIEW

Pawar Rohit and Rajeev R. Raje [19] used a dataset from tours, newspaper reviews and Twitter Tweets with Logistics Regression, Stochastics Gradient Descent and Multinomial Naive Bayes. Their outcomes shows that Logistics Regression outperforms SGD and MNB in all Hindi, Marathi and English. They further concluded that the performance and F1 score of all the ML algorithms improved by producing additional data using the technique of data integration.

Gaikwad et al. [9] released MOLD, the Marathi harmful dialect dataset, from about 2,500 commented tweets collected using the Twitter API and trained several individual language models on this dataset. Afterwards, they explored the most advanced multiple-language learning techniques to provide predictions to Marathi from Hindi, Bengali and English. They concluded that applying multi-language contextual word semantic vectors improved performance over monolingual models. Outcomes provided by their models confirmed that nearby interrelated languages link, Hindi and Marathi, provide an advantage in transfer learning experiments to improve performance.

Banerjee et al. [1] presented their recommended models for hurtful language identification and offensive speech detection in English, Hindi, Marathi and Code-mixed messages. They used XLM- Roberta and multilingual BERT (mBERT) for Marathi, Hindi and code-mixed classification on dataset sampled from Twitter and concluded that XLM-Roberta-large mostly achieves superior to diverse models rested on transformer imitations.

Chanda et al. [3] described their Fine-tuning Pre-Trained Transformer rest model for detecting harmful speech on Twitter in three dialects (i.e., Hindi, English and Marathi) and code-mixed (Hiindi-English) speech. The best-multilingual-cased model applied for the Marathi Language. Using the BERT implementation in the pytorch-transfomers library, they demonstrated their fine-tuned model. On the dataset from HASOC2021 for all classes, the full primarily trained model was trained and fed the outcome to a softmax level. The inaccuracy was backpropagated amongst the whole architecture; the model's primarily-trained weights were adjusted based on a new dataset, and it found that Marathi pre-processing did not yield as anticipated. It also mis-sorts specific NOT so long as hateful and offensive (HOF) compared to other languages.

Encoding the text into numeric vectors using the TF-IDF emphasize extraction approach on the Marathi dataset released by HASOC, Gajbhiye et al. [10] trained Random Forest Classifier and Logistic Regression and concluded that Random Forest with an accuracy of 0.77, performed comparatively better than the Logistic Regression approach.

Glazkova et al. [12] proposed a system for the Marathi dialect, which was rested on the Language- Agnostic BERT Sentence Embedding (LaBSE) primarily trained on 109 language texts, a data shared by HASOC 2021 and got an 88.08% F1 score.

Mandl et al. [18] expressed the significance of cede learning from a closely associated language while summarizing Hateful and Offensive Language Content Identification subtasks for English, Hindi, and Marathi Speeches using variations of transformer-rested algorithms presented at FIRE 21.

LSTM and CNN-rested models using and without using FastText vector representation assessed on 2021 HASOC Hindi and Marathi data. In addition, transformer-rested models, particularly interpretations of BERT like RoBERTa-base, mBERT and indicBERT, were used for Marathi to resemble by Velankar et al. [24]. CNN, LSTM, and BiLSTM elementary models were initially exercised with random text entrenching. The text vector representation likewise started using foremost-trained fast text vector representations by IndicNLP and afterwards utilised in static or trainable mode. The untrainable fast text vector representations seem more optimistic than random embedding and teachable fast text. Random and FastText vector representation and CNN and LSTM-rested models utilised for binary classification in Hindi and Marathi tasks. From these, untrained FastText setting + LSTM worked the best for Marathi. IndicBERT outperformed other imitations for Marathi.

Chanda et al. [2] used a system rested on fine-tuning the most recent converter models like XLM- RoBERTa, suitable for text processing through 100 different dialects and is exercised on considerably more exercise data than German BERT and BERT [5] to classify tweets in German, Marathi and Hindi-English Codemix and language. They concluded that primary-trained bidirectional encoder depiction utilizing transformers beat the conventional machine training models.

Chavan et al. [4] specified a thorough analysis of outcomes of many monolingual and multilingual BERT models like MuRIL: primarily trained on a vast multidialectal dataset overall many Indian dialects, MahaTweetBERT: primarily trained on L3CubeMahaTweetCorpus (848 thousands tokens) a large single language dataset including tweets expressed in the Marathi dialect[12], MahaTweetBERT- Hateful: exercised solely on abusive Marathi tweets pulled from the L3Cube-MahaTweet Dataset (23.7 thousands tokens), and MahaBERT: primarily-trained on L3Cube- Maha Dataset, a huge Marathi corpus including 725 thousands tokens, on the HASOC 2022/ MOLD v2 dataset and reflected shortcomings of these models.

In their approaches to creating additional data to enhance insufficient Resource-harmful Indic dialect identification, Das et al. [6] developed transformer-based classifier models for abusive speech identification using Python's Pytorch library.

They experimented on eight variant languages from openly available 14 resources under various settings - ELFI (Each language for itself), zero-shot learnedness, few-shot learnedness, cross-lingual learning, model cede, instance cede, etc. and concluded that the model's performance varies from language to language.

Velankar et al. [25] mentioned L3CubeMahaHate - a Tweet-based Marathi hateful text dataset and explored on LSTM, CNN and transformer-based single and multiple language variations of BERT as MahaBERT, IndicBERT, mBERT, and xlm-RoBERTa on binary and 4 class classification problems. They confirmed that untrainable fast word setting for CNN LSTM-rested models competed with BERT, exceeding indicBERT for both classes. The MahaBERT provided analysed outcomes, whereas MahaRoBERTa gave the significant accuracy of 4-class.

Gokhale et al. [13] introduced MahaTweetBERT and HindTweetBERT, the openly on-hand models BERT primarily exercised on Marathi and Hindi tweets individually. Their experiments indicated that hateful or non-hateful pre-training does not define the model performance from the observations of models pre-trained on biased tweets. They also saw that their monolingual models fare better than multilingual models like MuRIL, and the models pre-trained on all 40 million tweets performed better than the other, relatively smaller models. Their ensured results for both Marathi and Hindi languages concluded that observations are not language specific.

To classify the offensive and Non-offensive Marathi tweet text, Kalra et al. [15] used pre-trained BERT models like Multilingual [20]- pre-trained using 104 languages data, MuRIL [16]- BERT rested model domesticated over 17 Indian languages from data on Wikipedia *and* Distil [22]- a distilled version of BERT, with Fine-Tuning Classifier with conclusion of MuRIL outperformed among the three models.

Kumari Kirti and Jyoti Prakash Singh [17] experimented with HASOC 2022's code-mixed English, Hindi and Marathi languages data sets with Logistic Regression, Support Vector Machine, Multinomial Naïve Bayes, Decision Tree and Random Forest and presented their results on binary classification and 3 class classification using Random Forest, 4 class classification using Support Vector, over Marathi tweets. Some difficulties in pre-processing stop words faced them during Marathi text classification tasks, causing their F1 scores to be low.

Vani V Dikshithaand and Bharathi B [23] involved machine learning forecasting algorithms - Random forest (RF), Logistic Regression, Support Vector Machine(SVM) and k nearest neighbours (KNN) classifier algorithms besides score vectorized highlights to Marathi language tweets on hurtful dialect identification, auto-classification of hate varieties and offence taunt identification separately. Their results show that Random Forest foresees the labels for offensive language and offensive target identification more precisely than other classifier models. In contrast, the Logistic Regression classifier accurately identifies the automatic categorization of offence types.

A semi-supervised larger dataset called SeMOLD with 8000 plus Marathi occurrences was produced by Zampieri et al. [27] to predict the type and target of hurtful Social Media Posts in Marathi. For that, they tested a variety of machine learning models like SVC, BiLSTM, and CNN, including cutting-edge transformer models, investigated the recognition of abusive language using cross-lingual embeddings and relinquished learning because of it, they saw that from Hindi and English language, Marathi language performance gained on the three tests namely identification, classification, and target identification of offensive text.

Table1 summaries literature review of different approaches for identifying abusive and hateful speech on social media specifically in Marathi and other related languages.

Table. 1 Summary of approaches used for detecting offensive and hateful Marathi and other related Languages.

Language /Languages	**Dataset Resource**	**Approach/ Techniques**	**Best Result(s)**
Hindi and Marathi	Tour, newspaper reviews and Twitter Tweets	Multinomial Naive Bayes (MNB), Logistics Regression (LR), and Stochastics Gradient Descent (SGD)	F1 score 0.9412 using LR
Marathi from Bengali, Hindi, and English	Twitter	Cross-lingual learning	Macro F1 scores of 0.9401 from Hindi and of 0.9345 from Bengali
English, Marathi, Hindi and Code-mixed	Twitter	Multilingual BERT (mBERT) and XLM-Roberta	Macro F1 as 0.8756
English, Hindi, Marathi, and code mixed (English-Hindi)	Twitter	Transformer based model	Macro F1 score for Marathi 0.8545, for English 0.7976, for Hindi 0.7547 and for English-Hindi 0.6795
Marathi	Twitter	Logistic Regression (LR) and Random Forest (RF) Classifier.	F1 Score using LR 0.54 and 0.67 using RF
English and Marathi	Twitter	Transformer-based multilingual masked language model and a System based on the Language-Agnostic BERT Sentence Embedding (LaBSE)	F1 Score 0.8808 for Marathi and 0.8199 for English

English, Hindi, and Marathi	Twitter	Variants of of transformer-based algorithms	F1 measures 0.91, 0.78 and 0.83 for Marathi, Hindi and English
Marathi and Hindi	Twitter	Multi CNN model with IndicNLP FastText word embedding (Basic Model)	Macro F1 score for Marathi 0.842
		indicBERT	Macro F1 score for Marathi 0.869
Marathi and Hinglish Codemix and German	Twitter	Pre-trained bi-directional encoder representations using transformers	Macro F1: 0.935 for Marathi
Marathi	Twitter	MahaTweetsBERT	Macro F1: 98.43
Marathi and Code-Mixed-Language Hindi, English and German	Twitter	TF-IDF and Transformers-Based BERT-Variants	F1: 0.9233
code-mixed of English, Hindi and Marathi	HASOC (hasocfire.github.io) (Twitter)	Support Vector Machine, Logistic Regression, Multinomial Naïve Bayes, Decision Tree and Random Forest	F1: 0.92
Marathi	Twitter	Random forest (RF), Support Vector Machine(SVM), Logistic Regression, and k nearest neighbours (KNN)	Macro F1: 0.9745
Hindi and English language, Marathi	Twitter	SVC, BiLSTM, CNN and Transformers	Macro F1 0.85

5. THREAT AND ABUSIVE LANGUAGE IDENTIFICATION TOOLS

Number of tools that can be used for Threat and abusive Marathi Language detection as specified in Table 2 are freely available.

Most of these tools are transformer based and can be classified as Monolingual, multilingual and cross-lingual.

Table 2: HASOC in Marathi Identification tools

HASOC Identification tools	**Universal Resource Locator**
XLM-R	https://github.com/TharinduDR/MOLD/tree/master/experiments
XLM-RoBERTa (large-sized model)	https://huggingface.co/xlm-roberta-large
xlm-roberta-base	https://huggingface.co/xlm-roberta-base
KNN, RF, Logistic Regression & SVM	https://github.com/dikshu-02/HASOC2022-task3
ai4bharat/indic-bert	https://huggingface.co/ai4bharat/indic-bert
l3cube-pune/marathi-bert	https://huggingface.co/l3cube-pune/marathi-bert
l3cube-pune/marathi-albert	https://huggingface.co/l3cube-pune/marathi-albert
l3cube-pune/marathi-roberta	https://huggingface.co/l3cube-pune/marathi-roberta
l3cube-pune/marathi-bert-v2	https://huggingface.co/l3cube-pune/marathi-bert-v2
l3cube-pune/marathi-albert-v2	https://huggingface.co/l3cube-pune/marathi-albert-v2
l3cube-pune/mahahate-multi-roberta	https://huggingface.co/l3cube-pune/mahahate-multi-roberta
mahaNLP	https://pypi.org/project/mahaNLP/

6. THREAT AND ABUSIVE LANGUAGE IDENTIFICATION TOOLS APPROACHES

RoBERTa (Robustly Optimised BERT), BERT (Bidirectional Encoder Representations from Transformers), and ALBERT (A Lite BERT) are all models rested on converter. The transformer framework involves multi-self-immersion layers that capture contextual information from input text. ALBERT (A Lite BERT) is a transformer-based linguistic model that orients to minimise the multiple model parameters while maintaining or improving performance. ALBERT achieves this by applying various techniques of parameter reduction, like factorised vector representation parameterisation, betray-layer sharing attributes, and self-supervised sentence ordering prediction.

Single and multi-language modelling target trained model is XLM-RoBERTa [5]. Text stream samples from Single and 100 languages are collected, and the model formed to detect disguised words in the input.

Main differences between XLM-RoBERTa (large-sized model) and XLM-RoBERTa-base lie in their size, training data coverage, performance, and computational requirements. The large-sized model offers higher capacity and improved performance, but at the cost of increased computational demands.

Commonly used machine learning algorithms KNN, RF, Logistic Regression and SVM, as applied in sentiment analysis, can also be applied for hate and offensive language identification. Their interpretation relies on the size and quality of the dataset under training, feature representation techniques, and hyperparameter tuning [26].

Primarily trained exclusively on 12 primary Indian languages model is multilingual ALBERT IndicBERT. It is primarily exercised on our new single language corpus of nearby 9 thousand million tokens and assessed on different tasks afterwards. In contrast to other multi-language models (mBERT, XLM-R, etc.) IndicBERT has much fewer parameters, and it also achieves performance better than these models.

To achieve the most recent performance in multi-lingual understanding, the XLM-R model based on RoBERTa [8] uses self-supervised training techniques used with other languages [5] by training in one language.

A multilingual ALBERT (A Lite BERT) model pre-training with 12 important Indian languages and 9 billion token monolingual corpora called IndicBERT performs better than other multilanguage models (mBERT, XLM-R, etc.) while having far fewer parameters [14].

MahaBERT, MahaRoBERTa, and MahaAlBERT are Marathi language Indic BERT models that use L3Cube-MahaCorpus and other freely accessible Marathi monolingual datasets. Marathi BERT is a fine-tuned multilingual BERT (bert-base-multilingual-cased), Marathi RoBERTa is a multilingual RoBERTa (XLM-RoBERTa-base), and Marathi AlBERT is a multilingual ALBERT-based model [8].

On L3Cube-MahaHate fine-tuned MahaHate-multi-RoBERTa (Marathi Hate speech identification) is a MahaRoBERTa (l3cube-Pune/Marathi-RoBERTa) a tweet-rested Marathi language hateful speech detection dataset model. The dataset contains hate, offensive, profane, and not [25] four-class model labelled data.

In general, all these approaches used either transformer-based models or machine learning models that provides results based on limited corpus of Marathi Language collected from social media like twitter. Existing techniques and tools do not consider nuances, slags and cultural references in Marathi, as it is a rich language with several variations in dialect. This comparative study on various available tools and techniques highlight the result that more work is needed to be performed for Marathi language to gain best results by considering nuances in Marathi Language.

7. CONCLUSION

This paper has described a comparative study of Threat and Abusive Marathi Language detection tools and currently available techniques. These tools and techniques provide cogent results for high-recourse languages like English. Nevertheless, for regional low-recourse languages, especially Marathi, performance is based upon pretraining on similar types of language like Hindi, availability of corpus and word embedding techniques. Though work has been started recently on Hate and Offensive Dialects Identification in the Marathi speech, more tools and techniques are further needed due to the agglutinative nature of the language. Idea behind these new tools could be, rather than using transformer-based model, if models are trained by using more resources of reginal language like Marathi, more safety against such threaten and abusive content can be achieved on social media to keep it clean for the users of all ages.

8. REFERENCES

[1] Banerjee et al., “Exploring Transformer Based Models to Identify Hate Speech and Offensive Content in English and Indo-Aryan Languages.” arXiv, November 27, 2021. http://arxiv.org/abs/2111.13974.

[2] Chanda, Supriya, Sacchit D Sheth, and Sukomal Pal. “Coarse and Fine-Grained Conversational Hate Speech and Offensive Content Identification in Code-Mixed Languages Using Fine-Tuned Multilingual Embedding,” https://ceur-ws.org/Vol-3395/T7-3.pdf FIRE’22: Forum for Information Retrieval Evaluation, December 9-13, 2022, India

[3] Chanda et al., “Fine-Tuning Pre-Trained Transformer Based Model for Hate Speech and Offensive Content Identification in English, Indo-Aryan and Code-Mixed (English-Hindi) Languages,” Fire (2021). https://ceur-ws.org/Vol-3159/T1-44.pdf

[4] Chavan et al., “A Twitter BERT Approach for Offensive Language Detection in Marathi.” arXiv, December 20, 2022. http://arxiv.org/abs/2212.10039.

[5] Conneau et al., “Unsupervised cross-lingual representation learning at scale”, arXiv preprint arXiv:1911.02116 (2019).

[6] Das, Mithun, Somnath Banerjee, and Animesh Mukherjee. “Data Bootstrapping Approaches to Improve Low Resource Abusive Language Detection for Indic Languages.” arXiv, April 26, 2022. http://arxiv.org/abs/2204.12543.

[7] Devlin et al., “BERT: Pre-Training of Deep Bidirectional Transformers for Language Understanding.” arXiv, May 24, 2019. http://arxiv.org/abs/1810.04805.

[8] Joshi Raviraj, “L3Cube-MahaCorpus and MahaBERT: Marathi Monolingual Corpus, Marathi BERT Language Models, and Resources” arXiv:2202.01159 [cs.CL].

https://doi.org/10.48550/arXiv.2202.01159

[9] Gaikwad et al., "Cross-Lingual Offensive Language Identification for Low Resource Languages: The Case of Marathi." arXiv, September 8, 2021. http://arxiv.org/abs/2109.03552.

[10] Gajbhiye et al., "Machine Learning Models for Hate Speech Identification in Marathi Language," Fire (2021) https://ceur-ws.org/Vol-3159/T1-37.pdf

[11] Gamal et al., "Intelligent Multi-Lingual Cyber-Hate Detection in Online Social Networks: Taxonomy, Approaches, Datasets, and Open Challenges." Big Data and Cognitive Computing 7, no. 2 (March 24, 2023): 58. https://doi.org/10.3390/bdcc7020058.

[12] Glazkova et al., "Fine-Tuning of Pre-Trained Transformers for Hate, Offensive, and Profane Content Detection in English and Marathi," arXiv:2110.12687 [cs.CL],2021 https://doi.org/10.48550/arXiv.2110.12687

[13] Gokhale et al., "Spread Love Not Hate: Undermining the Importance of Hateful Pre-Training for Hate Speech Detection." arXiv, December 11, 2022. http://arxiv.org/abs/2210.04267.

[14] Kakwani et al., "IndicNLPSuite: Monolingual Corpora, Evaluation Benchmarks and Pre-trained Multilingual Language Models for Indian Languages" (https://aclanthology.org/2020.findings-emnlp.445

[15] Kalra et al., "Hate Speech Detection in Marathi and Code-Mixed Languages Using TF-IDF and Transformers-Based BERT-Variants," FIRE 2022: Forum for Information Retrieval Evaluation, December 9-13, 2022, India. https://ceur-ws.org/Vol-3395/T7-13.pdf

[16] Khanuja et al., Muril: Multilingual representations for indian languages, arXiv preprint arXiv:2103.10730 (2021).

[17] Kumari Kirti, and Jyoti Prakash Singh. "Machine Learning Approach for Hate Speech and Offensive Content Identification in English and Indo Aryan Code-Mixed Languages," Forum for Information Retrieval Evaluation, December 9-13, 2022, India https://ceur-ws.org/Vol-3395/T7-10.pdf

[18] Mandl et al. "Overview of the HASOC Subtrack at FIRE 2021: Hate Speech and Offensive Content Identification in English and Indo-Aryan Languages and Conversational Hate Speech." Proceedings of the 13th Annual Meeting of the Forum for Information Retrieval Evaluation (2021): n. pag.CEUR Workshop Proceedings (CEUR-WS.org) https://arxiv.org/abs/2112.09301

[19] Pawar Rohit, and Rajeev R. Raje. "Multilingual Cyberbullying Detection System." In 2019 IEEE International Conference on Electro Information Technology (EIT), 040–044. Brookings, SD, USA: IEEE, 2019. https://doi.org/10.1109/EIT.2019.8833846.

[20] Pires T., Schlinger E. and Garrette D., How multilingual is multilingual bert?, arXiv preprint arXiv:1906.01502 (2019).

[21] Ranasinghe et al., "Overview of the HASOC Subtrack at FIRE 2022: Offensive Language Identification in Marathi." arXiv, November 18, 2022. http://arxiv.org/abs/2211.10163.

[22] Sanh V. et al, DistilBERT, a distilled version of bert: smaller, faster, cheaper and lighter, arXiv preprint arXiv:1910.01108 (2019).

[23] Vani V Dikshithaand and Bharathi B "Hate Speech and Offensive Content Identification in Multiple Languages Using Machine Learning Algorithms," (FIRE). CEUR-WS. org, 2022 - ceur-ws.org

[24] Velankar et al., "Hate and Offensive Speech Detection in Hindi and Marathi." arXiv, November 1, 2021. http://arxiv.org/abs/2110.12200.

[25] Velankar et al., "L3Cube-MahaHate: A Tweet-Based Marathi Hate Speech Detection Dataset and BERT Models." arXiv, May 22, 2022. http://arxiv.org/abs/2203.13778.

[26] Velankar et al., "Mono vs Multilingual BERT for Hate Speech Detection and Text Classification: A Case Study in Marathi," 13739:121–28, 2023. https://doi.org/10.1007/978-3-031-20650-4_10.

[27] Zampieri et al., "Predicting the Type and Target of Offensive Social Media Posts in Marathi." Social Network Analysis and Mining 12, no. 1 (December 2022): 77. https://doi.org/10.1007/s13278-022-00906-8.