Multilingual Hate speech and Offensive language detection in English, Hindi, and Marathi languages

Adaikkan Kalaivani[1,2], Durairaj Thenmozhi[3]

[1]Department of Information and Communication Engineering, Anna University, Chennai
[2]Research Centre, Department of CSE, Sri Sivasubramaniya Nadar College of Engineering, Kalavakkam, TamilNadu
[3]Department of CSE, Sri Sivasubramaniya Nadar College of Engineering, Kalavakkam, TamilNadu

Abstract

Hate speech and offensive language are phenomena that spread with the rising popularity of social media forums. Automatic detection of such content is crucial for predicting conflicts among social communities and blocking inappropriate content from social media forums. This paper aims to describe our team SSN_NLP_MLRG submission to HASOC 2021: Hate speech and offensive language detection in English and Indo-Aryan language, where we explore different models to perform the subtask1 includes subtask A: To detect the comments is Hate speech and offensive (HOF) or NOT and subtask B: To classify the HOF comments into profanity (PRFN), Hate speech (HATE), Offensive (OFFN) in English, Hindi language and subtask A in Marathi language. The experiments cover different learning techniques that include machine learning, transfer learning, and Multilingual pre-trained models. Our best models are Roberta for English subtask A, BERT for English subtask B, and MBERT for the Hindi subtask A, Hindi subtask B, and Marathi subtask A. Our team achieved the macro-averaged F1 scores of 0.7919, 0.7320, 0.8223, 0.6242, and 0.5110 in the English subtask A, Hindi subtask A, Marathi subtask A, English subtask B, and Hindi subtask B, respectively.

Keywords

Transfer learning, Code-Mixed language, Machine learning, Language modeling, Low-resource language

1. Introduction

Social media is a vast online communication forum that enables the public to express themselves easily, at times, anonymously. While expressing their opinion oneself is a right of humans that is cherished, inducing and spreading offensive content towards another social community is an abuse of this liberty [1, 2]. Therefore, social media forums and other means of online communication platforms have begun to play a larger role in hate and offensive crimes. Many online social media forums such as Twitter, Facebook, Instagram, and YouTube consider hate speech and offensive content harmful and have the policy to remove such content. Due to the societal concern and how widespread offensive content is becoming on the Internet, there is a strong motivation to detect hate speech and offensive content in social media forums. Hate speech[1] defines the attacks against someone or group community, based on these attributes as

FIRE 2021, Forum for Information Retrieval Evaluation, December 13-17, 2021, India
kalaiwind@gmail.com, kalaivania@ssn.edu.in, (A. Kalaivani); theni_d@ssn.edu.in (D. Thenmozhi)
0000-0002-1497-5605 (A. Kalaivani); 0000-0003-0681-6628 (D. Thenmozhi)

[1]https://www.dictionary.com/browse/hate-speech

race, gender, ethnicity, religion, sexual orientation, age, physical or mental disability, and others. Offensive content[2] is a language that could seriously offend an individual or group based on their age, religious or political beliefs, marital or parental status, sexual orientation, physical features, national origin, or disability.

Hindi is an Indo-Aryan language with the official languages of India and spoken chiefly in North India. Marathi is an Indo-Aryan language spoken predominantly by Marathi people of the Maharashtra state in India and official and co-official language in the Maharashtra and Goa states of Western India. Code-mixed language is a phenomenon that combines one or more languages and also native language written in roman script. The detection and categorization of hate speech and offensive language in the indirect comments [3] of the code-mixed are challenging tasks not only in the English language. Therefore, there is an open research area in the field of a code-mixed multilingual community such as Hindi, Marathi languages, etc.

The HASOC 2019 [4] and 2020 competitions aim to train the systems capable of detecting hate speech and offensive content in social media forums for the English, Hindi, and German languages. In the HASOC 2019[3], the organizers offered sub-tasks A, B, C for English, Hindi, and sub-tasks A, B for German languages. Furthermore, whether the content is NOT or HOF (sub-task A), what are the characteristics of the HOF content (sub-task B), and who is the target of the HOF message (sub-task C). In the HASOC 2020[4], they organized the two tasks for English, German, and Hindi Languages, namely subtask A: To identify the given comment is HOF or NOT and subtask B: To categorize the HOF comment into offensive, hate, and profanity.

This paper presents our approaches to HASOC-2021. We have participated in subtask1 shared task consisting of English subtask A, English subtask B, Hindi subtask A, Hindi subtask B, and Marathi subtask A. The goal of subtask A is to identify and detect the social media comments are hate speech and offensive (HOF) or NOT. The subtask B aims to categorize the characteristics of the HOF content into hate, offensive, and Profanity. We used the machine learning algorithms, BERT, MBERT, ALBERT, RoBERTa, DistilBERT model with ktrain library, ULMFiT to adapt and fine-tune the system. We used the NLTK library for pre-processing the training set and testing set for all three languages. The paper outlines as follows. The survey of relevant works describes in Section 2. The details of the experiment data and technique of our models are in Sections 3 and 4. Section 5 describes the analysis part of the experiment results. Finally, Section 6 shows the concluded work and discusses further work.

2. Related work

The OffensEval 2019 [5] is the shared task to identify the offensive content in the English language. Most of the teams employed BERT with different parameters to detect offensive language. In the SemEval 2019 shared task, the researchers used machine learning approaches, bi-directional LSTM models to classify offensive language [6]. Mostly, the researchers have adapted and fine-tuned the BERT, GPT, and ULMFiT models for detecting offensive language in the shared task. The OffensEval 2020 [7, 8] is the Shared Task on Multilingual Offensive

[2]https://www.lawinsider.com/dictionary/offensive-content
[3]https://hasocfire.github.io/hasoc/2019/dataset.html
[4]https://hasocfire.github.io/hasoc/2020/dataset.html

Language Identification for the English, Greek, Danish, Turkish, Arabic languages. Most teams used con-textualized Transformers, ELMo embeddings, BERT, RoBERTa, and the multilingual mBERT to detect and categorize the offensive language in five different languages.

For the low-resource language, the authors used the cross-lingual data augmentation technique for the input context [9]. In the Gemeval2018 shared task [10], They obtained the optimal solution by using the maximum entropy meta-level classifier model to identify the micro-posts of offensive language in the German language. In the HatEval [11], the top team used the SVM model to detect the Twitter comments is hate speech against women and immigrants in multilingual language. The majority of teams used deep learning techniques in that LSTM model to detect hate speech and offensive language in the shared task of HASOC 2019.

In the HASOC 2020, the best-performed teams have used the variants of the BERT transformers model [12, 13] to identify and categorize the hate speech and offensive language in the English, German, and Hindi (code-mixed) languages. From the observation, most of the researchers were used machine learning techniques, deep learning techniques, variation of pre-trained transformer models to detect hate speech and offensive language. We are still dealing with the issue of recognizing offensive content in low-resource languages, as well as the problem of handling an imbalanced dataset in various code-mixed languages. This problem opens new research in different low-resource languages other than English. HASOC 2021[5] shared task organizers provide the resource for the English, Hindi (code-mixed), and Marathi languages [14].

3. Experiment data

This section presents the task description, data pre-processing techniques in the shared task HASOC 2021.

3.1. Data description

The organizers offer two shared tasks, namely subtask1 and subtask2. In subtask1, they provided the HASOC 2021 dataset of the English, Hindi (code-mixed), and Marathi languages. In subtask2, the organizers provided the Hindi code-mixed conversation dataset. Our team SSN_NLP_MLRG participated in the subtask1 for all three languages. Table 1 presents the annotated tweets for the English, Marathi, Hindi HASOC 2021 dataset. For training and testing the system, the English dataset has 3832 and 1281 posts. The Hindi code-mixed dataset contains 4594 posts, 1532 posts for training and testing the system. The Marathi code-mixed dataset consists of 1863 posts for the train system and 625 comments for testing the model system. Table 2 shows the statistics of the dataset for all three languages.

3.2. Task description

The shared task of HASOC 2021 is to identify the hate speech and offensive content in English and Indo-Aryan languages [15, 16]. The English language includes two tasks, namely subtasks

[5]https://hasocfire.github.io/hasoc/2021/dataset.html

Table 1
Sample annotated comments - HASOC2021

Comments	Label/Subtask A	Label/Subtask B
found the little bastard now the fun begins	**HOF**	**PRFN**
my first time seeing report about this is very heart breaking	NOT	NONE
technically that is still turning back the clock dick head	**HOF**	**OFFN**
india has got the worst finance minister and health minister ever	**HOF**	**HATE**

Table 2
Train dataset of HASOC 2021

Language	Label/Subtask A	No of Comments	Label/Subtask B	No of Comments	Total
English	HOF	2491	HATE	680	**3832**
English			OFFN	619	
English			PRFN	1192	
English	NOT	1341	NONE	1341	
Hindi	HOF	1433	HATE	566	**4594**
Hindi			OFFN	654	
Hindi			PRFN	213	
Hindi	NOT	3161	NONE	3161	
Marathi	HOF	663	-	-	**1863**
Marathi	NOT	1200	-	-	

A and B as same as Hindi language. The Marathi language offers one task, namely subtask A. Subtask A is a binary text classification task that focuses the systems able to classify the given social media comments into two classes, namely, HOF and NOT.

Hate and Offensive content (HOF): The social media comments contain harassment, profane, insults, threatening words.

Non-Hate and Offensive (NOT): The social media comments do not include hate and offensive content.

Subtask B is a multi-text classification that focuses the systems able to classify the given online comments into three classes, namely HATE, OFFN, PRFN.

Hate speech (HATE): The social media comments which contain hate words.

Offensive (OFFN): The social media posts which contain offensive content.

Profane (PRFN): The social media posts contain profanity words.

3.3. Data pre-processing

The data pre-processing is to clean the social media comments from the unnecessary noisy content is present in the given dataset and transform it into a coherent form, which can be portable for English, Hindi code-mixed, and Marathi languages. We used the NLTK[6] for data cleaning, data duplication from the HASOC 2021 dataset [17]. First, we remove @ symbol with a string denoted as user-id because it does not have any meaningful expressions. Next, we remove

[6]https://www.nltk.org/

Table 3
Validation accuracy of the BERT-based and language models

Model	English A	English B	Hindi A	Hindi B	Marathi A
BERT	0.84	0.66	-	-	-
ALBERT	0.81	0.66	-	-	-
MBERT	-	-	0.79	0.69	0.88
DistilBERT	0.82	-	-	-	-
ULMFiT	0.75	-	-	-	-
RoBERTa	0.83	-	-	-	-

the hashtag with a text as the user's name because it affects the performance of our model. The example of data pre-processing like "@AjeebBharti @BeraJaykrishna @khan_nainam @Policy @Twitter Prove it !!! What evidence you have bloody hell prove kar" and after that "prove it what evidence you have bloody hell prove kar". After that, we removed the punctuation, numerals, symbols, URLs, and emojis and then converted the upper case text into small case text. Finally, we replaced the misspelling offense words and string with * into appropriate matched words presented in the collected vocabulary words.

4. Methodology

This section presents the experimental analysis of the various methods used for the validation process.

4.1. BERT-based model and Language Model

We have experimented with various pre-trained models, namely BERT uncased, DistilBERT base-uncased, ALBERT (Albert-base-v2), RoBERTa base, ULMFiT language modeling, and Machine learning techniques for subtask A of English language. We used the BERT, ALBERT pre-trained models and machine learning techniques for subtask B of the English language. For the Hindi language subtask A and B, We used the multi-cased BERT transformers to adapt and fine-tune the system to classify the hate speech and offensive content from the given dataset. For the Marathi language subtask A, we used the multi-cased BERT (MBERT) for the binary text classification task. For the validation process of the system, we take 25% of the data from the training dataset for the three languages. We used the above-mentioned pre-trained models with the ktrain[7] library that is useful to build the system using machine learning, neural network, and deep learning techniques. We have analyzed the training system to set the various batch size to 6, 32 and learning rates as 2e-5, 3e-5, and the epochs to 6, 7, 9, and 10. We used the ULMFiT[18] framework in that Average-SGD Weight-Dropped LSTM (AWD-LSTM) architecture model to predict the hate speech and offensive content and their characteristics for the English language dataset. Table 3 presents the validation results for the BERT-based models of the three languages.

[7]https://pypi.org/project/ktrain/

4.2. Machine Learning Techniques

For the machine learning techniques, we have conducted two experiments.

4.2.1. Experiment 1

In the first experiment, we have used the following models, namely support vector machine classifier (SVM), Naive Bayes classifier (NB), random forest classifier (RF), and Extreme gradient boosting ensemble classifier (XGB), and used to predict the hate speech and offensive content in the given English dataset. We used the sci-kit learn library for the implementation of the machine learning classifiers. For using Term frequency-inverse document frequency (TF-IDF) vectorization, we extracted the Ngram, character level, word-level features from the given dataset. For using the sklearn CountVectorizer, we build vocabulary for known words and also tokenize the collected data. FastText is the pre-trained vector for 157 languages trained on Common crawl and Wikipedia. We used the FastText pre-trained word embedding vectors for the English language, namely Wikipedia Tamil vectors (wiki.ta.vec).

4.2.2. Experiment 2

In the second experiment, we used the Gensim library for vector embeddings. Gensim is the fastest library for training the system of vector embedding. We experiment with the logistic regression, Multinomial Naive Bayes (NB), Random forest, and Linear Support vector machine (SVC) models to predict the system. We have utilized the genism for pre-processing and lemmatized the training dataset for this experiment 2. We have utilized the word cloud for categorized the training dataset. We have used a Document to vector transformer (Doc2vec) and text to TFIDF transformer for extracted the features by using the genism model. Table 4 presents the validation results for the machine learning models of the English language.

Finally, we have used the MBERT model to predict the hate speech and offensive content and got a macro F1-score of 0.8223 with the epochs to 10 and the learning rate as 2e-5 for the Marathi subtask A. We got macro F1-scores of 0.7320, 0.511 of the Hindi subtask A, Hindi subtask B with the epochs to 10, and the learning rate as 2e-5 for the MBERT model. For English subtask A, We got a macro F1 score of 0.7919 for the RoBERTa model with the 07 epochs and the learning rate as 3e-5. We got a macro F1 score of 0.624 with the 09 epochs and the learning rate as 2e-5 for the BERT model of the English subtask B.

5. Result Analysis

This section presents the evaluation of our model. For instance, we used the evaluation metrics like precision, recall, macro-averaged F1-score. We have submitted our best model after comparing the performance of our methods for English, Marathi, and Hindi code-mixed languages. The HASOC 2021 organizers provided the test data for English subtask A, English subtask B, Hindi subtask A, Hindi subtask B, and Marathi subtask A. From the performance of the validation system, the RoBERTa model achieved an accuracy of 0.83 and Precision, Recall, and macro F1-score of 0.81, 0.80, and 0.80, when compared with the performance of the other machine

Table 4
Validation accuracy of the machine learning models

Model	English A	English B
Experiment 1:		
NB Count vector	0.74	-
NB WordLevel TF-IDF	0.70	-
NB N-gram	0.68	-
NB CharLevel	0.69	-
SVM N-gram	0.68	-
RF, Count vector	0.73	-
RF, Word level	0.72	-
XGB, Count vector	0.72	-
XGB, Word level	0.72	-
XGB, CharLevel	0.73	-
Experiment 2:		
LR doc2vec	0.65	0.36
RF doc2vec	0.64	0.36
XGB doc2vec	0.64	0.35
NB doc2vec	0.48	0.21
SVM doc2vec	0.65	0.36
LR TFIDF	0.74	0.61
RF TFIDF	0.75	0.62
XGB TFIDF	0.69	0.62
NB TFIDF	0.70	0.58
SVM TFIDF	0.73	0.59

learning approaches and pre-trained language models. The F1-score for the Not-offensive comments and hate speech offensive comments for the RoBERTa model is 0.74, 0.87 respectively. Therefore, the RoBERTa model performs well than other models for the English subtask A. The accuracy of the English subtask B is 0.66, and the F1-score for the OFFN, HATE, PRFN, and NONE comments for the BERT model are 0.54, 0.73, 0.41, and 0.77.

From the observation, the BERT model performs well than other approaches of machine learning techniques for the English subtask B. For the Hindi language, the MBERT model achieved an accuracy of 0.79 for subtask A, 0.69 for subtask B, and F1-score of HOF and NOT comments are 0.65, 0.86 for the subtask A, and the F1-score for the OFFN, HATE, PRFN, and NONE comments for the subtask B task are 0.83, 0.19, 0.40 and 0.43 respectively. For Marathi Language, MBERT achieved an accuracy of 0.88, the macro F1-score is 0.87, and the F1-score of HOF and NOT comments are 0.91 and 0.83. We adapted and fine-tuned the BERT-based model to build and predict the data and its characteristics for all the languages. Our team submitted three runs for the English subtask A and one runs for the subtasks of the other languages.

The final results[8] of our team for the three languages are present in Table 5. Our team SSN_NLP_MLRG submission got the 19^{th}, 12^{th}, 25^{th}, 9^{th}, 19^{th}rank in the shared task for English subtask A, English subtask B, Hindi subtask A, Hindi subtask B, Marathi subtask A respectively.

[8]https://hasocfire.github.io/hasoc/2021/results.html

Table 5
Final results for the three languages

Language	Model	Accuracy	Precision	Recall	Macro F1
English A	RoBERTa	0.80	0.80	0.78	**0.79**
English A	ALBERT	0.79	0.78	0.77	0.77
English A	BERT	0.80	0.80	0.77	0.78
English B	ALBERT	0.65	0.61	0.60	0.60
English B	BERT	0.66	0.62	0.62	**0.62**
Hindi A	MBERT	0.77	0.74	0.72	**0.73**
Hindi B	MBERT	0.70	0.49	0.53	**0.51**
Marathi A	MBERT	0.84	0.82	0.82	**0.82**

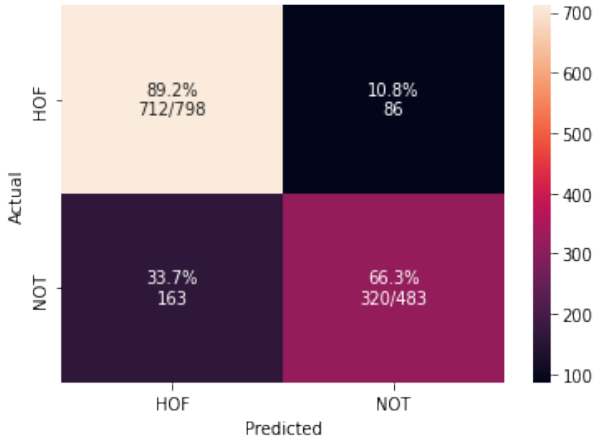

Figure 1: Confusion matrix of English subtask A - RoBERTa

We classify the performance of the model for all the three languages by using the confusion matrix are presented in the Figure 1 for English subtask A for the RoBERTa model, Figure 2 for the English subtask B for the BERT model, Figure 3 for the Hindi subtask A for the MBERT model, Figure 4 for the Hindi subtask B for the MBERT model, Figure 5 for the Marathi subtask A for the MBERT model. From the confusion matrix, we noticed that many test cases were classified as HOF comments by the RoBERTa model for the English subtask A. For English A, the Precision, Recall, and F1-score for the HOF and NOT comments are 0.81, 0.89, 0.85, and 0.79, 0.66, 0.62 respectively. For English B, the F1-score for the OFFN, HATE, PRFN, and NONE comments are 0.72, 0.46, 0.75, and 0.56. For Hindi A, the Precision, Recall, and F1-score for the HOF and NOT comments are 0.81, 0.87, 0.84, and 0.69, 0.58, 0.63 respectively. For Hindi B, the F1-score for the OFFN, HATE, PRFN, and NONE comments are 0.83, 0.40, 0.50, and 0.31. For Marathi A, the Precision, Recall, and F1-score for the HOF and NOT comments are 0.89, 0.88, 0.88, and 0.75, 0.77, 0.76 respectively. Overall, the hate speech and offensive comments perform well by Bert-based models for all three languages.

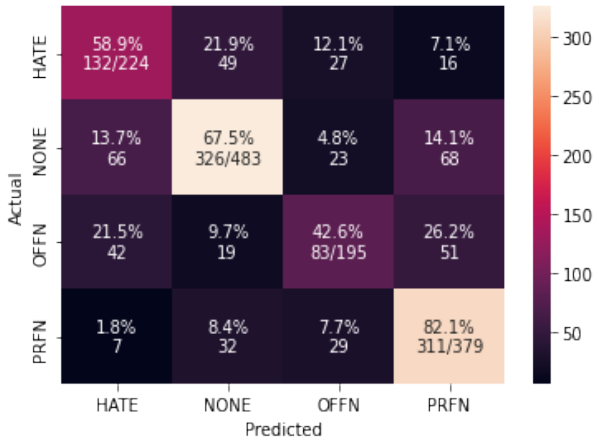

Figure 2: Confusion matrix of English subtask B - BERT

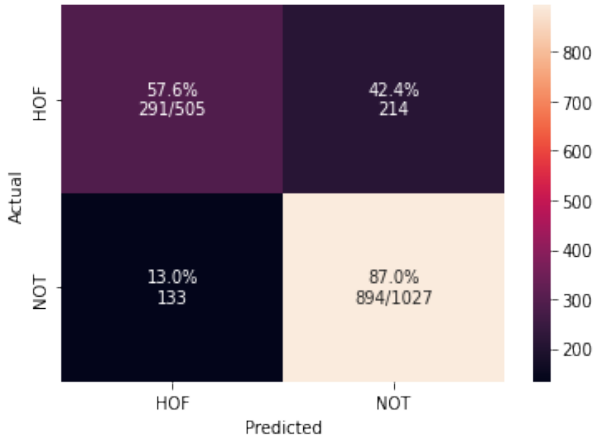

Figure 3: Confusion matrix of Hindi subtask A - MBERT

6. Conclusion

This paper presents the team submitted runs for the hate speech and offensive language identification for the HASOC 2021 subtask1 shared task in the Forum for Information Retrieval Evaluation (FIRE) 2021. We experimented with different approaches such as machine learning techniques, pre-trained BERT-based models. The results show the RoBERTa models perform well than the other BERT-based models and the machine learning approaches for the English subtask A. The BERT uncased performs well in the English subtask B. MBERT performs well in Hindi subtask A, Hindi subtask B, and Marathi subtask A. Based on the evaluation, the Overall BERT-based model performs well for the three languages. Our team submission had a macro F1-score of 0.8223, for the Marathi subtask A, macro F1-score of 0.7320, 0.511 for the Hindi subtask A and Hindi subtask B code-mixed language, and macro F1-score of 0.7919, 0.624 for the

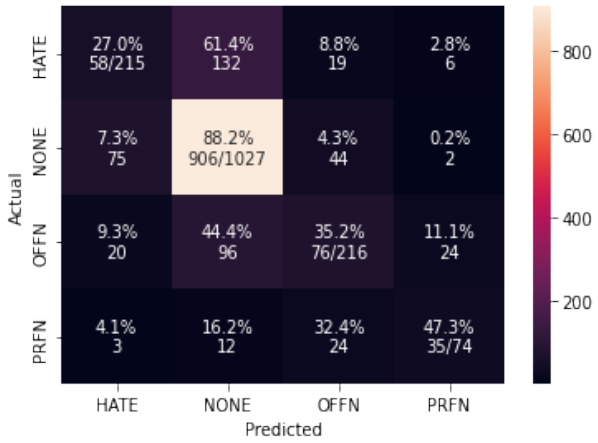

Figure 4: Confusion matrix of Hindi subtask B - MBERT

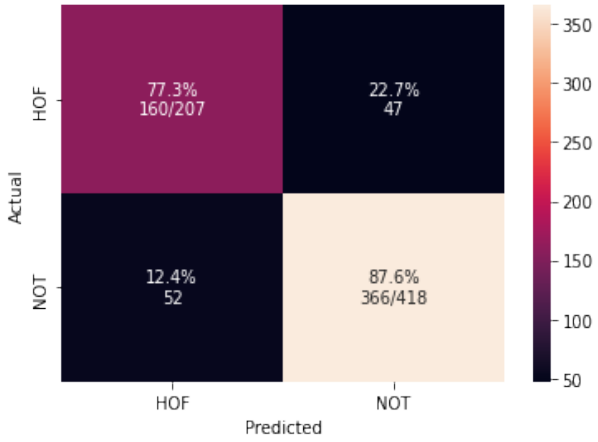

Figure 5: Confusion matrix of Marathi subtask A - MBERT

English subtask A and English subtask B. For future work, we will handle the sarcastic feature and imbalanced dataset to avoid misclassification and extend this work into other low-resource languages.

References

[1] S. MacAvaney, H.-R. Yao, E. Yang, K. Russell, N. Goharian, O. Frieder, Hate speech detection: Challenges and solutions, PloS one 14 (2019) e0221152.

[2] A. Kalaivani, D. Thenmozhi, Sentimental Analysis using Deep Learning Techniques, International Journal of Recent Technology and Engineering (IJRTE) 7 (2019) 600–606.

[3] A. Kalaivani, D. Thenmozhi, Sarcasm Identification and Detection in Conversion Context using BERT, in: Proceedings of the Second Workshop on Figurative Language Processing,

Association for Computational Linguistics, Online, 2020, pp. 72–76. URL: https://www.aclweb.org/anthology/2020.figlang-1.10. doi:10.18653/v1/2020.figlang-1.10.

[4] T. Mandl, S. Modha, P. Majumder, D. Patel, M. Dave, C. Mandlia, A. Patel, Overview of the HASOC Track at FIRE 2019: Hate Speech and Offensive Content Identification in Indo-European Languages, in: Proceedings of the 11th Forum for Information Retrieval Evaluation, FIRE '19, Association for Computing Machinery, New York, NY, USA, 2019, p. 14–17. URL: https://doi.org/10.1145/3368567.3368584. doi:10.1145/3368567.3368584.

[5] M. Zampieri, S. Malmasi, P. Nakov, S. Rosenthal, N. Farra, R. Kumar, SemEval-2019 Task 6: Identifying and Categorizing Offensive Language in Social Media (OffensEval), 2019. arXiv:1903.08983.

[6] D. Thenmozhi, B. Senthil Kumar, S. Sharavanan, A. Chandrabose, SSN_NLP at SemEval-2019 Task 6: Offensive Language Identification in Social Media using Traditional and Deep Machine Learning Approaches, in: Proceedings of the 13th International Workshop on Semantic Evaluation, Association for Computational Linguistics, Minneapolis, Minnesota, USA, 2019, pp. 739–744. URL: https://www.aclweb.org/anthology/S19-2130. doi:10.18653/v1/S19-2130.

[7] M. Zampieri, P. Nakov, S. Rosenthal, P. Atanasova, G. Karadzhov, H. Mubarak, L. Derczynski, Z. Pitenis, Çağrı Çöltekin, SemEval-2020 Task 12: Multilingual Offensive Language Identification in Social Media (OffensEval 2020), 2020. arXiv:2006.07235.

[8] A. Kalaivani, D. Thenmozhi, SSN_NLP_MLRG at SemEval-2020 Task 12: Offensive Language Identification in English, Danish, Greek Using BERT and Machine Learning Approach, in: Proceedings of the Fourteenth Workshop on Semantic Evaluation, International Committee for Computational Linguistics, Barcelona (online), 2020, pp. 2161–2170. URL: https://aclanthology.org/2020.semeval-1.287.

[9] J. Singh, B. McCann, N. S. Keskar, C. Xiong, R. Socher, XLDA: Cross-Lingual Data Augmentation for Natural Language Inference and Question Answering, CoRR abs/1905.11471 (2019). URL: http://arxiv.org/abs/1905.11471. arXiv:1905.11471.

[10] M. Wiegand, M. Siegel, J. Ruppenhofer, Overview of the GermEval 2018 Shared Task on the Identification of Offensive Language, in: In Proceedings of GermEval, 2018.

[11] V. Basile, C. Bosco, E. Fersini, D. Nozza, V. Patti, F. M. Rangel Pardo, P. Rosso, M. Sanguinetti, SemEval-2019 Task 5: Multilingual Detection of Hate Speech Against Immigrants and Women in Twitter, in: Proceedings of the 13th International Workshop on Semantic Evaluation, Association for Computational Linguistics, Minneapolis, Minnesota, USA, 2019, pp. 54–63. URL: https://www.aclweb.org/anthology/S19-2007. doi:10.18653/v1/S19-2007.

[12] A. Kalaivani, D. Thenmozhi, SSN_NLP_MLRG@HASOC-FIRE2020: Multilingual Hate Speech and Offensive Content Detection in Indo-European Languages using ALBERT, in: P. Mehta, T. Mandl, P. Majumder, M. Mitra (Eds.), Working Notes of FIRE 2020 - Forum for Information Retrieval Evaluation, Hyderabad, India, December 16-20, 2020, volume 2826 of *CEUR Workshop Proceedings*, CEUR-WS.org, 2020, pp. 188–194. URL: http://ceur-ws.org/Vol-2826/T2-12.pdf.

[13] T. Mandl, S. Modha, A. Kumar M, B. R. Chakravarthi, Overview of the HASOC Track at FIRE 2020: Hate Speech and Offensive Language Identification in Tamil, Malayalam, Hindi, English and German, in: Forum for Information Retrieval Evaluation, FIRE 2020, Association for Computing Machinery, New York, NY, USA, 2020, p. 29–32. URL: https:

//doi.org/10.1145/3441501.3441517. doi:10.1145/3441501.3441517.
[14] S. Gaikwad, T. Ranasinghe, M. Zampieri, C. M. Homan, Cross-lingual Offensive Language Identification for Low Resource Languages: The Case of Marathi, in: Proceedings of RANLP, 2021.
[15] S. Modha, T. Mandl, G. K. Shahi, H. Madhu, S. Satapara, T. Ranasinghe, M. Zampieri, Overview of the HASOC Subtrack at FIRE 2021: Hate Speech and Offensive Content Identification in English and Indo-Aryan Languages and Conversational Hate Speech, in: FIRE 2021: Forum for Information Retrieval Evaluation, Virtual Event, 13th-17th December 2021, ACM, 2021.
[16] T. Mandl, S. Modha, G. K. Shahi, H. Madhu, S. Satapara, P. Majumder, J. Schäfer, T. Ranasinghe, M. Zampieri, D. Nandini, A. K. Jaiswal, Overview of the HASOC subtrack at FIRE 2021: Hate Speech and Offensive Content Identification in English and Indo-Aryan Languages, in: Working Notes of FIRE 2021 - Forum for Information Retrieval Evaluation, CEUR, 2021. URL: http://ceur-ws.org/.
[17] S. Bird, E. Klein, E. Loper, Natural language processing with Python: Analyzing text with the natural language toolkit, "O'Reilly Media, Inc.", 2009.
[18] A. Kalaivani, D. Thenmozhi, SSN_NLP_MLRG@ Dravidian-CodeMix-FIRE2020: Sentiment Code-Mixed Text Classification in Tamil and Malayalam using ULMFiT., in: FIRE (Working Notes), 2020, pp. 528–534.