Overview of the shared task on Offensive Language Identification in Dravidian code-Mixed Languages

Sripriya N[1,*,†], Bharathi Raja Chakravarthi[2,†], Thenmozhi Durairaj[1,†], Bharathi B[1,†], Prasanna Kumar Kumaresan[2,†], Subalalitha C N[3,†], Anusha M D[4,†], Parameshwar R Hegde[4,†] and Deepthi Vikram[4,†]

[1]Sri Sivasubramaniya Nadar College of Engineering, Tamil Nadu, India

[2]University of Galway, Galway, Ireland

[3]SRM Institute of Science And Technology, Tamil Nadu, India

[4]Yenepoya Institute of Arts Science Commerce and Management,Mangalore, India

Abstract

Offensive language detection is a critical task in natural language processing, particularly in the context of online discourse, where harmful content can spread rapidly. Identifying offensive language is challenging due to the varied ways in which offense is conveyed, including subtle linguistic cues, code-mixing, and cultural context. Code-mixing is a prevalent phenomenon in a multilingual community, and the code-mixed texts are sometimes written in non-native scripts. Systems trained on monolingual data fail on code-mixed data due to the complexity of code-switching at different linguistic levels in the text.This shared task presents a gold-standard dataset for offensive language detection in Tamil, Malayalam, Kannada, and Tulu, enabling researchers to develop robust classification models.13 teams actively participated and developed systems to identify offensive content in this shared task. This work summarizes the various techniques used by the competing teams. Further, the performance analysis of the all systems was carried out using macro-F1 score and their rankings were reported.

Keywords

Offensive Identification, corpus Creation, Classification, Code-Mixing, Dravidian Languages, shared task

1. Introduction

The detection of offensive language has become a pivotal task within the field of Natural Language Processing (NLP), especially in the contemporary digital landscape characterized by the widespread dissemination of user-generated content across social media platforms [1]. Online environments such as YouTube and Instagram facilitate millions of daily interactions encompassing topics related to entertainment, politics, social issues, and personal viewpoints. While these platforms promote open communication, they simultaneously function as environments conducive to the propagation of offensive material, including abusive language, hate speech, cyberbullying, misogyny, targeted harassment and sarcasm [2][3]. The ramifications of such conduct are extensive, leading to psychological distress, perpetuating social inequalities—particularly affecting vulnerable populations such as women—and compromising the safety and well-being of online communities.

Growing concerns regarding online abuse have stimulated increased scholarly attention toward the development of automated systems designed to detect and mitigate harmful content. Nonetheless, the task of offensive language detection remains complex due to the intricate linguistic manifestations of offensive expressions. Such content may be overt or covert, frequently articulated through sarcasm,

metaphorical language, cultural allusions, or contextual nuances that pose significant interpretative challenges for computational models. These difficulties are further exacerbated in multilingual contexts, where individuals often engage in code-mixing, i.e., alternating between two or more languages within a single utterance or discourse.

Code-mixing is commonly found in South Indian linguistic groups, where languages such as Tamil, Malayalam, Kannada, and Tulu frequently blend with English during casual conversations. The situation becomes more complicated when these languages are represented in non-native scripts or Romanized formats, leading to significant disparities in spelling, morphology, and grammatical structures. Conventional NLP systems, which are usually trained on monolingual and well-formed text, find it challenging to cope with such variability, resulting in subpar performance on actual code-mixed data. Furthermore, languages like Kannada and Tulu are under-resourced, and the lack of annotated datasets hinders the creation of effective offensive language identification models.

To fill these gaps, the FIRE 2025 Global Offensive Language Detection Challenge aims to detect offensive content in code-mixed Dravidian languages, particularly Tamil-English, Malayalam-English, Kannada-English, and Tulu-English. Common Challenge provides a benchmark dataset compiled from YouTube comments across entertainment, news, and sociopolitical domains. Participants are asked to categorize comments into categories such as offensive, non-aggressive, targeted insults, and non-targeted (individual/group) insults.

The challenge is expected to support applications such as automated content moderation on social media platforms, help law enforcement identify harmful behavior, and help organizations monitor public opinion while ensuring respectful communication. Beyond direct applications, the overall goal is to advance multilingual and multilingual NLP research by encouraging the development of models that can understand different linguistic patterns, scriptural variations, and culturally specific expressions.

This highlights the need for comprehensive datasets and methodologies that effectively generalize across languages, scripts, and communities and help create safer and more equitable online environments.

The rest of this document is organized as follows. Section 2 reviews related studies. Section 3 describes the problem and dataset. Section 4 presents the methodology adopted by the participating teams with the results and Section 5 concludes the article.

2. Related Work

Offensive speech is communication that offends, or hurts feelings of others. On social media sites, these kinds of content are inciting hatred, encouraging violence, and targeting specific people or groups based on their identification. Henceforth, it is essential to filter and censor violence in order to promote healthy online communication and safeguard vulnerable populations [4]. The HASOC-track at FIRE2019 [5] focused on hatespeech and offensive content detection in Indo-European languages, while HASOC at FIRE2020 [6] extended the focus to HateSpeech and OLI in Tamil, Malayalam, Hindi, English, and German. Furthermore, HASOC Dravidian-CodeMix at FIRE2021 [7] targetedcode-mixed Tamil and Malayalam,whereas DravidianLangTechatEACL2021 [8] [9] addressed OLI in code-mixedDravidianlanguages(Tamil-English, Malayalam-English,and Kannada-English).These tasks have significantly contributed to the creation of annotated datasets,evaluation metrics, and baselinemodels for offensive language identification in Indian and Dravidian languages,fostering inclusive and safer digital communication environments.

A growing number of researchers are using transformer-based models, such as BERT [10], IndicBERT [11] to enhance semantic comprehension. These models are particularly well-suited to handling the complexities of Indian languages since they were pretrained on a variety of multilingual datasets. Despite these advancements, Tulu and Kannada remain relatively understudied in comparison to Tamil and Malayalam [12] [13].

Table 1

Dataset

Languages	Labels	Train	Dev	Test	Total
Kannada	Not_offensive	3,544	426	427	7,772
	not-Kannada	1,522	191	185	
	Offensive_Targeted_Insult_Individual	487	66	75	
	Offensive_Targeted_Insult_Group	329	45	44	
	Offensive_Untargetede	212	33	33	
	Offensive_Targeted_Insult_Other	123	16	14	
Malayalam	Not_offensive	14,153	1,779	1,765	20,010
	not-malayalam	1,287	163	157	
	Offensive_Targeted_Insult_Individual	239	24	27	
	Offensive_Untargetede	191	20	29	
	Offensive_Targeted_Insult_Group	140	13	23	
Tamil	Not_offensive	25,425	3,193	3,190	43,919
	Offensive_Untargetede	2,906	356	368	
	Offensive_Targeted_Insult_Individual	2,557	307	315	
	Offensive_Targeted_Insult_Group	2,343	295	288	
	not-Tamil	1,454	172	160	
	Offensive_Targeted_Insult_Other	454	65	71	
Tulu	not offensive	1,261	270	270	3,845
	not tulu	726	156	155	
	offensive untargeted	462	99	99	
	offensive targeted	243	52	52	

3. Task Description and Datasets

This shared task presents a corpus for offensive language identification of code-mixed text in Dravidian languages (Tamil-English, Malayalam-English, Kannada-English and Tulu-English). This task is further complicated in low-resource languages where limited annotated datasets exist for offensive speech detection. This task presents a gold-standard dataset for offensive language detection in Tamil, Malayalam, Kanada and Tulu, enabling researchers to develop robust classification models.

The primary goal of this shared task is to build and evaluate systems that can automatically classify social media text into these four categories. Participants will be provided with training, development, and test datasets to develop their models. Given the real-world class imbalance in offensive content, models must be designed to handle the skewed distribution of data effectively. As far as we know, this is the first shared task on offensive language detection in Tulu. By organizing this task, we aim to foster research in under-resourced languages, improve computational approaches for offense detection in multilingual and code-mixed settings, and contribute to the responsible use of AI in moderating harmful content online.

4. Dataset

The dataset consists of social media comments and posts that are categorized into four classes:

- Not Offensive (NO): Content without any offensive elements.
- Offensive Untargeted (OU): Offensive content that is not directed at a specific individual or entity.
- Offensive Targeted (OT): Direct attacks on an individual or group, including hate speech targeting a community, ethnicity, caste, or gender.
- Not Tamil/Not Malayalam/ Not Kanada/ Not Tulu (NT): Content that does not contain the corresponding language. Table 1 shows the distribution of the training set, validation set, and test set for this task for all the four languages.

Table 1 shows the distribution of the training set, validation set, and test set for this task for all the four languages.

5. Methodology and Results

A total of 13 teams participated in the Offensive Language Identification shared task at FIRE 2025, where the the participants submitted different runs/models and the performance of these models was measured using the metric called macro F1-score. The teams were ranked using the macro F1-score which is a metric trying to balance both the measures, precision and recall. This metric is good for evaluating the offensive identification models, because false positives and false negatives are well balanced by this measure. The models created by all the teams show promising capabilities in spite of the challenges posed by linguistic diversity, code-mixed, and social disparities.

The team **"CoreFour"** followed a comprehensive methodology that included data preprocessing, traditional machine learning, and transformer-based deep learning techniques. As a first step, various NLP-based preprocessing such as lowercasing, punctuation removal, and normalization of repeated characters were carriedout. To handle multilingual inconsistencies and reduce noise, a custom self-translation mapping was implemented to correct transliterated or code-mixed words into a more standardized form. For the classification models, traditional algorithms like Support Vector Machines (SVM) and Random Forest using TF-IDF features extracted at word, character, and subword levels were trained. Fine-tuning of the state-of-the-art multilingual transformer models such as mBERT, Distil-mBERT, IndicBERT, and XLM-RoBERTa to capture rich contextual semantics was performed. Finally, the outputs of these models were combined using ensemble strategies like soft voting and model blending to improve robustness and accuracy. This multi-faceted approach effectively addressed the challenges of linguistic diversity and label imbalance, resulting in more reliable and generalizable performance. This system has secured a top score of 0.778 for Malayalam-English and top 3 positions in the remaining three code-mixed languages.

"NLPFusion" [14] team have used different transformer model and also employed balanced weights. The proposed methodology makes use of Transfer Learning (TL) based Multilingual Bidirectional Encoder Representations from Transformers (mBERT) model and XLM-RoBERTa. These models achieved macro F1 scores of 0.465, 0.475, and 0.820, securing first rank in Tamil, Kannada, Tulu respectively and second rank in Malayalam with 0.774 score.

The team **"MUCS_Off"** [15] preprocessed the Datasets of Kannada, Malayalam, Tulu, and Tamil as the first step in the structured methodology of identifying offensive language in multiple Dravidian languages. Transforming text to lowercase and eliminating URLs, mentions, and other superfluous characters while keeping those unique to Indian scripts are examples of text preprocessing. In order to make the preprocessed text appropriate for neural network models, it is subsequently transformed into numerical sequences using a tokenizer and padded to guarantee consistent length. LSTM, CNN, and a hybrid model that combines the two are the three categories of deep learning models are constructed. CNNs extract local features, LSTMs capture sequential dependencies, and the hybrid model combines the two advantages.To avoid overfitting and enhance generalization, these models are trained with early stopping and learning rate reduction. These systems exhibited average performance compared to other teams in the shared task.

IREL@IIT-BHU [16] team have used fine-tuned pre-trained XLM-RoBERTa model with and without early stopping. Across the four languages, these systems consistently ranked within the top-6, achieving 4th place in Kannada and Tamil, 5th in Malayalam, and 7th in Tulu, thereby establishing multilingual transformers as a strong baseline for Dravidian code-mixed offensive language identification.

The **"DUCS"** team employed a comprehensive, two-pronged approach, systematically comparing classical machine learning techniques with a state-of-the-art transformer-based model. The entire workflow was encapsulated in a robust, end-to-end OffensiveLanguagePipeline for reproducibility. Initially, a strong baseline using several traditional models was established, including Logistic Regression, SVM, Random Forest, and Multinomial Naive Bayes. The text data underwent a careful preprocessing pipeline tailored for code-mixed Tamil-English content, which involved normalizing whitespace, replacing URLs, mentions, and hashtags with special tokens, and a unique step of converting emojis into textual representations For e.g., EMOJI_smile to retain their semantic value. Features for these models were generated using TF-IDF vectorization with both unigrams and bigrams to capture lexical patterns. The

core of our system is a fine-tuned transformer model, specifically google/muril-base-cased, chosen for its strong performance on Indian languages and code-mixed text. Recognizing the severe class imbalance in the dataset, we implemented two distinct strategies: for the baseline models, SMOTE (Synthetic Minority Over-sampling Technique) was used to balance the training data at the data level. For the more sophisticated transformer model, we implemented a custom WeightedTrainer class that applies class weights directly to the Cross-Entropy loss function, compelling the model to pay more attention to underrepresented offensive categories during training. The system was designed to evaluate both approaches and select the best-performing model for final predictions and finally showed a macro F1-score of 0.416.

In order to address offensive language detection in Tulu, Kannada, Tamil and Malayalam, the team **"YenLP_CS"** [17] used a multi-stage ensemble-based methodology that combined transformer-based deep learning techniques with conventional machine learning techniques. To establish a solid statistical baseline, the TF-IDF features have been extracted from the text and trained a group of classifiers, including Logistic Regression, XGBoost, MLP, SVC, and KNN, all of which were combined via a voting mechanism. They have integrated two transformer-based models to improve contextual understanding: (1) IndicBERT, where CLS-token embeddings were extracted and fed into a custom MLP classifier, and (2) XLM-Roberta, where an XGBoost classifier was trained using CLS-token embeddings. Finally, a hard majority voting strategy was used to fuse the predictions from the TF-IDF ensemble, IndicBERT+MLP, and XLM-Roberta+XGBoost. This proposed method was good in identifying offensive language content in Malayalam-mixed English text securing 0.75 macro-f1 score.

The **"Dravidian_decoders"** [18] team have adopted an ensemble-based machine learning approach to perform offensive language classification across two South Indian languages: Kannada and Tulu. The methodology began with thorough data preprocessing, which included text normalization and label encoding to make the datasets suitable for machine learning. For feature extraction, TF-IDF vectorization with both unigrams and bigrams was used to capture important contextual patterns in the text. Three lightweight yet effective classifiers—Linear Support Vector Machine (SVM), Logistic Regression, and Multinomial Naive Bayes—were trained separately on each language-specific dataset. These models were then combined using a majority voting ensemble technique to aggregate predictions and improve classification robustness. The ensemble model capitalizes on the strengths of each individual classifier, thereby improving generalization and handling of class imbalances. The models were evaluated using F1-score and realized that are yielding average performances securing 6th and 8th rank in Kanada and Tulu.

The **"Coreminds"** [19] team developed a multilingual Offensive language classification system for four low-resource Indian languages. The methodology involved collecting, cleaning, and preprocessing labeled datasets for each language, ensuring consistent formatting and removing noise. Labels were encoded using LabelEncoder to convert them into numeric format. Fine-tuned two powerful transformer-based language models were used for training: IndicBERTv2-MLM-only, which is trained on 12 Indian languages, and TwHIN-BERT-base, a BERT-based model pre-trained on Twitter data, which is particularly effective for handling social media text. These models were chosen based on the nature of the data and the language coverage. This approach ensures adaptability to noisy, real-world text in underrepresented Indian languages. This proposed model was performing well for Malayalam-mixed English securing 3rd rank and was showing average performance for rest of the languages.

The **"DravidianDefenders"** team adopted a traditional machine learning-based approach tailored individually for each language. The methodology involved extensive preprocessing to clean and normalize the code-mixed social media comments, including removal of URLs, special characters, emojis, and non-linguistic artifacts. For text representation, TF-IDF vectorization with unigrams and bigrams were used to effectively capture the linguistic patterns in code-mixed contexts. Different classifiers were used for each language: Logistic Regression for Tamil, Random Forest for Malayalam and Tulu, and Support Vector Machine (SVM) for Kannada. For Kannada, the model was further optimized using character n-gram TF-IDF features and class-weighted SVM with grid search. The models exhibited only average performance and obtained the ranks above 6 across all four languages.

The methodology used by the **"langTeam"** [20] for this multi-class text classification task involved

several key steps. Initially, the dataset was loaded and subjected to a comprehensive preprocessing pipeline. This included basic text cleaning, tokenization, stop word removal, stemming, and lemmatization. Subsequently, the preprocessed text data was transformed into numerical features using both TF-IDF and Count Vectorization techniques. These vectorizations were applied to various levels of processed text. Multiple classification models were then employed, including Logistic Regression, Naive Bayes, Support Vector Classifier, Decision Tree, Random Forest, and Gradient Boosting. A BiLSTM model was also implemented, utilizing Word2Vec embeddings. Each model was trained and evaluated on the different vectorized representations of the data. The predictions of the model with best performance was taken based on it's accuracy. The systems showed F1 score of 0.267,0.511, 0.77 for Tamil, Malayalam and Tulu language.

The **"Malayalam_lan_tech"** team employed a supervised learning pipeline using transformer-based sentence embeddings combined with classical machine learning classifiers for performing the offensive language identification in Malayalam. Initially, the data was preprocessed by cleaning and encoding the labels into numerical format. A lightweight, pre-trained multilingual transformer model was utilized to generate fixed-size sentence embeddings for the Malayalam code-mixed text. These embeddings served as features for training multiple classifiers, including K-Nearest Neighbors (KNN), Random Forest (RF), Support Vector Machine (SVM), and XGBoost. The model was tested against Malayalam language dataset and obtained a score of 0.14.

The systems developed by all the participated teams for all the four language code-mixed datasets were evaluated and ranked based on the F1 Score and are provided in Tables 2, 3, 4 and 5.

Table 2
Tamil-English Task-Ranklist

S.No.	Team Name	Run	mF1	Rank
1	NLPFusion [14]	run2	0.465	1
2	CoreFour	run3	0.461	2
3	MUCS_off [15]	run1	0.452	3
4	IREL@IIT-BHU [16]	run2	0.448	4
5	DUCS	run1	0.416	5
6	DravidianDefenders [21]	-	0.386	6
7	YenLP_CS [17]	run2	0.369	7
8	MUCS_DL [15]	run2	0.350	8
9	Coreminds [19]	-	0.308	9
10	langTeam [20]	-	0.267	10

Table 3
Malayalam-English Task-Ranklist

S.No.	Team Name	Run	mF1	Rank
1	CoreFour	run1	0.778	1
2	NLPFusion [14]	run1	0.774	2
3	YenLP_CS [17]	run3	0.750	3
4	MUCS_offensive [15]	run2	0.712	4
5	IREL@IIT-BHU [16]	run1	0.667	5
6	langTeam [20]	-	0.511	6
7	DravidianDefenders [21]	run1	0.430	7
8	MUCS_DL [15]	run3	0.350	8
9	Coreminds [19]	-	0.233	9
10	Malayalam_lan_tech	-	0.140	10

Table 4
Kannada-English Task-Ranklist

S.No.	Team Name	Run	mF1	Rank
1	NLPFusion [14]	run1	0.475	1
2	CoreFour	run3	0.434	2
3	Coreminds [19]	-	0.433	3
4	IREL@IIT-BHU [16]	run2	0.430	4
5	MUCS_Off [15]	run1	0.421	5
6	Dravidian_decoders [18]	run1	0.384	6
7	YenLP_CS [17]	run2	0.375	7
8	MUCS_DL [15]	run1	0.344	8

Table 5
Tulu-English Task-Ranklist

S.No.	Team Name	Run	mF1	Rank
1	NLPFusion [14]	1	0.820	1
2	MUCS_off [15]	3	0.790	2
3	CoreFour	1	0.780	3
4	Dravidian_decoders [18]	2	0.770	4
5	langTeam [20]	-	0.770	4
6	YenLP_CS [17]	2	0.730	5
7	IREL@IIT-BHU [16]	1	0.710	6
8	MUCS_DL [15]	3	0.610	7
9	DravidianDefenders [21]	-	0.550	8
10	Coreminds [19]	-	0.400	9

6. Conclusion

In this shared task,we have promoted the issue of detecting offensive language in code-mixed Dravidian languages, specifically Tamil-English,Malayalam-English, Kannada-English and Tulu-English. The research community is encouraged by this mission to investigate new methods for developing strong and trustworthy offensive language identification systems. Four datasets containing postings that had been scraped from social media platforms in Tamil, Malayalam, Kanada, and Tulu were provided. Thirteen teams took part and created systems employing a variety of techniques, such as transformer-based models and conventional machine learning models. The F1 score was used to assess and rank the outcomes according to the model's effectiveness. Future study in this area will benefit from the ideas that each team employed to develop their systems, which were also emphasized.

Acknowledgments

Declaration on Generative AI

In the course of preparing this manuscript, the author(s) employed the generative AI tool ChatGPT. Its use was limited to performing checks for grammar and spelling. Following this, the author(s) conducted a thorough review and revision of the text and assume full responsibility for the final published content.

References

[1] P. Fortuna, S. Nunes, A survey on automatic detection of hate speech in text, ACM Comput. Surv. 51 (2018). URL: https://doi.org/10.1145/3232676. doi:10.1145/3232676.

[2] B. R. Chakravarthi, S. N, B. B, N. K, T. Durairaj, R. Ponnusamy, P. K. Kumaresan, K. K. Ponnusamy, C. Rajkumar, Overview of sarcasm identification of dravidian languages in dravidiancodemix@ fire-2023, in: Forum of Information Retrieval and Evaluation FIRE - 2023, Goa, India, 2023.

[3] S. N, B. B, T. Durairaj, N. K, R. Ponnusamy, P. K. Kumaresan, K. K. Ponnusamy, C. Rajkumar, B. R. Chakravarthi, Overview of sarcasm identification of dravidian languages in dravidiancodemix@fire-2024, in: Forum of Information Retrieval and Evaluation FIRE - 2024, DAIICT , Gandhinagar, 2024.

[4] S. Sai, Y. Sharma, Towards offensive language identification for Dravidian languages, in: B. R. Chakravarthi, R. Priyadharshini, A. Kumar M, P. Krishnamurthy, E. Sherly (Eds.), Proceedings of the First Workshop on Speech and Language Technologies for Dravidian Languages, Association for Computational Linguistics, Kyiv, 2021, pp. 18–27. URL: aclanthology.org.

[5] T. Mandl, S. Modha, andetal. et al., 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, 2019.

[6] T. Mandl, S. J. Modha, M. Anandkumar, B. R. Chakravarthi, Overview of the hasoc track at fire 2020: Hate speech and offensive language identification in tamil, malayalam, hindi, english and german, Proceedings of the 12th Annual Meeting of the Forum for Information Retrieval Evaluation (2020). URL: https://api.semanticscholar.org/CorpusID:231628577.

[7] P. K. Kumaresan, Premjith, R. Sakuntharaj, S. Thavareesan, S. Navaneethakrishnan, A. K. Madasamy, B. R. Chakravarthi, J. P. McCrae, Findings of shared task on offensive language identification in tamil and malayalam, in: Proceedings of the 13th Annual Meeting of the Forum for Information Retrieval Evaluation, FIRE '21, Association for Computing Machinery, New York, NY, USA, 2022, p. 16–18. URL: https://doi.org/10.1145/3503162.3503179. doi:10.1145/3503162.3503179.

[8] B. R. Chakravarthi, R. Priyadharshini, N. Jose, T. Mandl, P. K. Kumaresan, R. Ponnusamy, J. P. McCrae, E. Sherly, et al., Findings of the shared task on offensive language identification in tamil, malayalam, and kannada, in: Proceedings of the first workshop on speech and language technologies for Dravidian languages, 2021, pp. 133–145.

[9] B. R. Chakravarthi, R. Priyadharshini, V. Muralidaran, N. Jose, S. Suryawanshi, E. Sherly, J. P. McCrae, Dravidiancodemix: Sentiment analysis and offensive language identification dataset for dravidian languages in code-mixed text, Language Resources and Evaluation 56 (2022) 765–806.

[10] J. Devlin, M.-W. Chang, K. Lee, K. Toutanova, BERT: Pre-training of deep bidirectional transformers for language understanding, in: J. Burstein, C. Doran, T. Solorio (Eds.), Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers), Association for Computational Linguistics, Minneapolis, Minnesota, 2019, pp. 4171–4186. URL: https://aclanthology.org/N19-1423/. doi:10.18653/v1/N19-1423.

[11] D. Kakwani, A. Kunchukuttan, S. Golla, G. N.C., A. Bhattacharyya, M. M. Khapra, P. Kumar, IndicNLPSuite: Monolingual corpora, evaluation benchmarks and pre-trained multilingual language models for Indian languages, in: T. Cohn, Y. He, Y. Liu (Eds.), Findings of the Association for Computational Linguistics: EMNLP 2020, Association for Computational Linguistics, Online, 2020, pp. 4948–4961. URL: https://aclanthology.org/2020.findings-emnlp.445/. doi:10.18653/v1/2020.findings-emnlp.445.

[12] B. R. Chakravarthi, M. B. Jagadeeshan, V. Palanikumar, R. Priyadharshini, Offensive language identification in dravidian languages using mpnet and cnn, International Journal of Information Management Data Insights 3 (2023) 100151. URL: https://www.sciencedirect.com/science/article/pii/S2667096822000945. doi:https://doi.org/10.1016/j.jjimei.2022.100151.

[13] A. M. D, D. Vikram, B. R. Chakravarthi, P. R. Hegde, Overcoming low-resource barriers in tulu: Neural models and corpus creation for offensive language identification, 2025. URL: https://arxiv.

org/abs/2508.11166. `arXiv:arXiv:2508.11166`.

[14] H. Asha, S. M, Amrithkala, M. Shazia, C. Sharal, Multilingual pretrained models for offensive language identification in dravidian code-mixed text, in: Forum of Information Retrieval and Evaluation FIRE - 2025, Varanasi, India, 2025.

[15] N. Rachana, S. Hosahalli, Lakshmaiah, Exploringclassicalmachinelearninganddeeplearning approachesforoffensivelanguageidentificationindravidian code-mixedtext, in: Forum of Information Retrieval and Evaluation FIRE - 2025, Varanasi, India, 2025.

[16] T. Krishna, C. Supriya, A. P. K, Irel@iit-bhu@dravidiancodemix 2025: Offensive language identification, in: Forum of Information Retrieval and Evaluation FIRE - 2025, Varanasi, India, 2025.

[17] A. Raksha, S. Rathnakara, Yenlp_cs@dravidiancodemix 2025: A trifusion model for offensive language detection in dravidian code-mixed text, in: Forum of Information Retrieval and Evaluation FIRE - 2025, Varanasi, India, 2025.

[18] S. P, A. S, A. V, D. J, Dravidiandecoders@ dravidiancodemix 2025: Offensive content classification in kannada–tulu code-mixed texts using classical machine learning, in: Forum of Information Retrieval and Evaluation FIRE - 2025, Varanasi, India, 2025.

[19] S. P, A. A, V, M. S, Arul, C. T, Coreminds@dravidiancodemix 2025: Comparative study of transformer-based models for offensive content detection in tamil, malayalam, kannada, and tulu code-mixed texts, in: Forum of Information Retrieval and Evaluation FIRE - 2025, Varanasi, India, 2025.

[20] S. Y, R. P, Saicharan, K. K, Revanth, S. D.V.L, Lang team@dravidiancodemix 2025: Offensive detect, in: Forum of Information Retrieval and Evaluation FIRE - 2025, Varanasi, India, 2025.

[21] S. R, S, S. U, S, S. M, S. P, S, Dravidiandefenders@dravidiancodemix 2025: Empirical analysis of classical machine learning approaches in tamil, malayalam, and tulu code-mixed offensive content classification, in: Forum of Information Retrieval and Evaluation FIRE - 2025, Varanasi, India, 2025.