[1]Brahma Naidu K
[2] Vishnuvardhan B
[3]Adi Narayana Reddy K

Enhancing Hate Speech Detection with Integrated Content-Based and Stylistic Features

Journal of Electrical Systems

Abstract: - The proliferation of harmful and unpleasant speech on community medium platforms has underscored the need for effective hate speech detection. While recent efforts have focused on refining pre-trained models, this study takes a novel approach by emphasizing the integration of content-based and stylistic features. Stylistic features, in particular, play a critical role in hatred speech detection. By capturing unique linguistic patterns and characteristics indicative of hateful or offensive language—beyond explicit content—these features enhance the discriminatory power of detection systems. In this research, exploration of combined utilization of SVM, XGBoost, and Random Forest algorithms on a comprehensive dataset. The results surpass existing methodologies, contributing to more effective identification and mitigation of problematic content online.

Keywords: Hate speech, Machine learning, TF-IDF, Stylistic features, Word Embeddings.

I. Introduction

Hate communication refers to any type of communication which diminishes with target individuals or gatherings based on properties like contest, color, society, sexual orientation, gender, religion, nationality or additional distinguishing qualities [1]. With huge amount of content generated by user on platforms like Twitter, addressing the identification and prevention of hate speech has become crucial, especially in combating misogyny and xenophobia.

Our goal is to identify potential sources of hate speech on Twitter as a first step towards curbing its spread among online users. Twitter guidelines prohibit tweets from making threats or engaging in harassment based on ethnicity, gender, religion, or other attributes. Similarly, YouTube restricts content promoting violence or hostility towards specific individuals or groups, including age, caste, and disabilities [2]. The surge in online information sharing highlights the pressing requirement for automated hatred speech detection.

Given a diverse hate speech laws globally, the challenge lies in defining boundaries in cyberspace and addressing the gap in manual oversight by internet administrators. This proliferation poses a significant challenge to policymakers and researchers. Modern developments in NLP technologies encompass spurred research into automating hate speech detection in textual content.

To address the specified problems, the distasteful content classification and hated speech Indo-Aryan Language and English introduced three communal tasks across multiple languages [3]. In this study, we specifically address the first task in the English language. The challenge involves classifying content into two distinct categories: HOF that is Hate or Offensive, which includes content containing hatred speech, unpleasant language, or vulgarity, and NOT that is Non Hate or Offensive, which refers to content that, is free from hatred speech, profanity and offensive language.

In this paper, we follow the following structure: First, we discuss previous approaches in the related work section (Section 2). Next, Section 3 outlines the backdrop and outlines our projected method. During the section 4, presents the results and perform a comparison analysis. Lastly, Section 5 contains conclusion of our work.

II. Literature Survey

Detecting hate speech presents a significant research challenge, as evidenced by the existing literature employing various methodologies, including dictionary-based approaches [4], distributional semantics [5], and neural network architectures [6, 7]. However, a significant amount of the proposed research has primarily focused on detestation speech detection with English. In contrast, here has been limited scholarly attention to other foreign languages [8, 9, 10, 11, 12] in addition to the complexities with the code switching of text [13, 14]. Even though with the significant influence of provincial little resource language on online hatred speech, current

[1]*Corresponding author:Department of Artificial Intelligence and Data Science, Faculty of Science and Technology, (IcfaiTech), ICFAI Foundation for Higher Education(IFHE),Hyderabad-501203, India&Research Scholar, JNT University, Hyderabad Email: brahmanaidu@ifheindia.org.
[2]Senior Professor, Dept. of CSE, JNT University, Hyderabad.
[3]Department of CSE, Faculty of Science and Technology, (IcfaiTech), ICFAI Foundation for Higher Education (IFHE), Hyderabad-501203,

field remnants moderately uncharted. Current investigations have begun to explore the utility of transformers [15] and author profiling using graph neural networks [16].

Throughout the history of hate speech detection research, a variety of strategies have been explored. Kwok et al, [17] initially experiment by means of a basic bag of words approach but encountered challenges through lofty false positive counts. Incorporating additional NLP components like Ngram methods and part of speech tags [18] enhanced the presentation of hatred speech uncovering models. Additionally, techniques combining TF-IDF with Support Vector Machines (SVM) showed promising results [19].

The introduction of word embedding's like GloVe [20] and FastText [21] has significantly advanced hate speech detection by mapping text into a latent space, surpassing traditional BOW and lexical methods. Recurrent Neural Networks (RNNs), especially a single layer Bidirectional LSTM models using the fast text embedding's, have demonstrated success in discerning hate speech, exemplified by the captivating methodology in 2020 HASOC contest for the Hindi [22]. Similarly, LSTM architectures with GloVe embedding's have proven effective for English [23], a trend also supported by Mohtaj et al. [24] using character-based LSTMs.

More recently, transformer models leveraging self-attention mechanisms [15] and variants like BERT [25], trained on extensive corpora, have exhibited superior performance over traditional RNNs across various NLP tasks. BERT-like models are well-known for their transfer learning capabilities [26].

Regardless of the considerable amount of the study on the hatred speech recognition, studies focused on low resource language are still limited. The logistic regression using LASER embedding's shown better performance than BERT models [27], underscoring the necessitate for additional accurate and efficient multi language based models. Multilingual language models like XLM-Roberta [28] have gained prominence, and region-specific low-resource language models such as MuRIL [29], SinBERT [30], BanglaBERT [31], XLMIndic [32] and IndicBERT [33] have emerged. Researchers in [35] conducted a comprehensive study on monolingual versus multilingual model performance for cross-lingual hatred speech recognition.

Preceding edition of HASOC [36, 37] encompass seen substantial efforts to enhance presentation in low resource language like Marathi [38] and Hindi [39], among others. In the upcoming section, we will delve into our approach, which leverages multiple multi lingual models for identifying hatred speech, and provide a comprehensive comparative analysis of these models against alternative methodologies.

III. METHODOLOGY

In this, different machine learning based concepts used in the proposed methodology is discussed

A. Background:

SVM:

The primary objective of SVM be toward determine a optimal hyper plane that segregates distinct variations within the feature space. In binary classification, this hyper plane acts as the decision boundary that maximizes the margin (or distance) between the closest data points from various variations, known as support vectors

Random Forest Algorithm:

It is a highly adaptable and potent the ensemble learning technique utilized with both regression and classification problems. It falls under the category of bagging ensemble techniques and is renowned for its resilience and adaptability when dealing with intricate datasets. Let's delve deeper into the workings of the Random Forest algorithm

XGBoost:

XGBoost is an ensemble learning technique that constructs a series of weak learners (often decision trees) and merges them to form a robust learner. It operates within a gradient boosting framework, progressively building models to rectify errors from preceding iterations.

B. Data Representation:

TF-IDF:

We have employed the method TF-IDF that is term frequency with inverse document frequency for assessing of topics of every document depends on its contained words. TF-IDF assigns weights to words, measuring relevance rather than mere frequency. Specifically, word counts are replaced with TF-IDF scores across the entire corpus. Initially, TF-IDF calculates how often words appear within a specific document (term frequency). However, common words such as "and,""or," or "the" which appear frequently across all documents are systematically discounted (inverse-document frequency). The rationale is that words appearing in many documents provide less value in distinguishing any particular document. This process aims to highlight only the frequently occurring and distinctive words as indicators. We generate uni-grams, bi-grams, and tri-grams with a given Twitter post's bag of words illustration. To accommodate potential variations in the content extent between the training with test datasets, these features are represented using TF-IDF values.

Word Embedding's:

Word2Vec be the widely used method to generate word embeddings in NLP, originally evolved by programmers in Google. This method employs the superficial neural network model for representing vocabulary as opaque vectors within a continuous vector space, where vocabulary with like meaning or context are situated closer together.

Global Vector be the representative of Word Representation that is GloVe is a different well-liked method for generating vocabulary embeddings in NLP, developed by researchers in Stanford University. GloVe construct word embeddings with utilization of global word to word co occurrence information derived with a given text corpus. Unlike Word2Vec, GloVe operates by factorizing the word co-occurrence matrix to effectively capture statistical relationships between words.

GloVe embeddings excel in capturing both semantic and syntactic similarities among words, making them valuable for plentiful NLP applications like word analogy, sentiment investigation, and machine translation. These pre-trained embeddings are often integrated into deep learning models to enhance performance across different tasks. GloVe is preferred for its capability to produce high-quality word embeddings grounded in comprehensive corpus statistics, enabling precise representations of word meanings and contexts in textual data.

POS Tagging

POS tagging represents a foundational task within NLP, involving the assignment of a specific part of speech to every word in a given text. These POS tags serve to classify the syntactic role of each word within a sentence, encompassing categories such as nouns, verbs, adjectives, adverbs, prepositions, conjunctions, pronouns, and others. POS tagging be the foundational to understanding the grammatical structure of text and is indispensable for building sophisticated NLP systems capable of nuanced language processing and analysis.

Stylistic Features

Stylistic features are essential in hate speech detection as they capture distinct linguistic patterns and characteristics that signal hateful or offensive language, going beyond explicit content. These features focus on elements like writing style, tone, and expression, complementing semantic and contextual features to improve detection accuracy.

Offensive Language and Profanity Detection: Identifying explicit profanity, hate speech terms, slurs, and derogatory language based on frequency and intensity, which are strong indicators of hate speech.

Sentence Structure Analysis: Examining sentence length, complexity, and grammatical irregularities, as hate speech often displays simpler structures or unconventional grammar.

Punctuation and Symbol Usage: Analyzing excessive utilization of punctuation marks like exclamation marks, question marks, and symbols (e.g., emojis, emoticons) to gauge emotional intensity or aggression.

Capitalization Identification: Detecting words or phrases written in all capital letters, which can convey shouting or emphasis commonly associated with hateful expression.

Repetition Detection: Noticing repeated words, phrases, or patterns, which can emphasize hateful rhetoric or propaganda.

C. Anticipated System Architecture

The anticipated system architecture is presented in Fig. 1. Initially preprocessing the applied on the tweets. Preprocessing is vital in hate speech detection to ready text data for analysis and model training. As part of preprocessing, eliminating unnecessary characters like special symbols, emoji's, URLs, and HTML tags that do not aid content analysis. After this, transfer text to lower case for ensuring uniformity in words (e.g., "HATE" and "hate" are treated equally). Expanded contractions (e.g., "can't" to "cannot") and replace common abbreviations (e.g., "u" to "you"). Finally Identify and remove common stop word vocabulary (e.g., "the", "is", "and") that lack significant meaning.

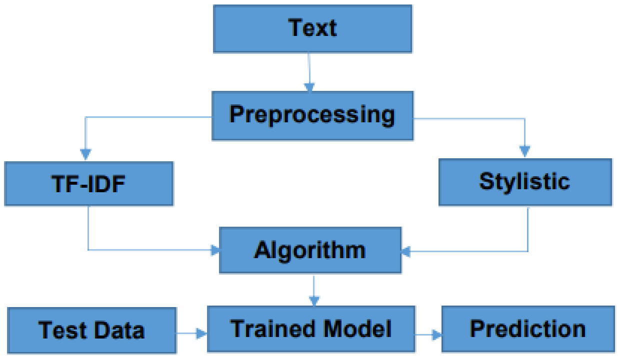

Fig. 1 Proposed Architecture

IV. Implementation and Result Analysis

A. Datasets

HASOC offers a platform and a data challenge to facilitate multilingual research aimed at identifying hatred speech and unpleasant data. The details of datasets year wise presented in Table. 1, Table 2 and Table 3. This proposed task focusing on for identifying hatred speech and unpleasant language in Hindi, English, and Marathi. The proposed sub task A involves coarse grained binary categorization, requiring participating system to categorize tweets into two distinct categories: HOF that is Hate or Offensive, which includes content containing hatred speech, unpleasant language, or vulgarity, and NOT that is Non Hate or Offensive, which refers to content that, is free from hatred speech, profanity and offensive language.

They consist of normal statement or content. Utterances that are considered "normal" and non-offensive should not be labeled as such, as they may be element of adolescence language or other linguistic register. As part of our research we considered only English tweets and combined HASOC 2019, HASOC2020 and HASOC 2021

Table 1 A Sample of Tweets from HASOC 2019

Class	No. Samples
HOF	2549
NOT	4456

Table 2 A Sample of Tweets from HASOC 2020

Class	No. Samples
HOF	1856
NOT	1852

Table 3 A Sample of Tweets from HASOC 2021

Class	No. Samples
HOF	2501
NOT	1342

B. Implementation Details:

We have implemented SVM, Random Forest, XGBoost algorithms on content-based features, content-based plus stylistic features, context based word embedding’s and content based word embedding’s plus stylistic features. We have considered F1-Score as the performance measure.

Result Analysis and Discussion

In this section, we have showcased the outcomes achieved using our proposed models. Following that, we have assessed the performance of our developed system in comparison to existing systems.

i) Content-based TFIDF

We have combined TF-IDF and POS tagging scores and implemented SVM, Random Forest and XGBoost algorithms. The f1-scores are presented in Table 4. We got 78.1 as the highest f1-score for the SVM algorithm

ii) Content-based TFIDF Plus Stylistic

We have combined TF-IDF, POS tagging and stylistic features and implemented SVM, Random Forest and XGBoost algorithms. The f1-scores are presented in Table 5. We got 80.3 as the highest f1-score for the SVM algorithm

iii) Word Embeddings (Word2Vec)

We have created 300 dimension word embeddings using word2vec and implemented SVM, Random Forest and XGBoost algorithms. The f1-scores are presented in Table 6. We got 78.8 as the highest f1-score for the SVM algorithm

iv) Word Embeddings (GloVe)

We have created 200 dimension word embeddings using GloVe and implemented SVM, Random Forest and XGBoost algorithms. The f1-scores are presented in Table 7. We got 79.2 as the highest f1-score for the SVM algorithm.

Table 4 F1-score(TF-IDF)

Algorithm	F1-score
SVM	78.1
Random Forest	77.8
XGBoost	77.9

Table 5 F1-score(TF-IDF Plus Stylistic)

Algorithm	F1-score
SVM	80.3
Random Forest	78.6
XGBoost	78.8

Table 6 F1-score(Word2Vec)

Algorithm	F1-score
SVM	78.8
Random Forest	77.9
XGBoost	78.4

Table 7 F1-score(GloVe)

Algorithm	F1-score
SVM	79.2
Random Forest	78.7
XGBoost	78.74

Comparison with Other Works

We have conducted a comparison between our proposed approach and the results provided by the participating systems in the HASOC competition. The methodologies employed for this comparison are detailed below.

Table 8. Result Comparison

Algorithm	Methodology	Macro F1 score
BERT[]	FastText	80.24
BERT[]	BERT Encoder, Character Encoder, Hate words Encoder	80.18
Magnified TIDS[]	TF-IDF	80.13
BiLSTM	Fusion of TF-IDF, BERTweets	80.06
Proposed work	Combined TF-IDF and Stylistic	80.3

TF-IDF assigns weights to words, measuring relevance rather than mere frequency. Stylistic features play a crucial role in hate speech detection by capturing unique linguistic patterns and characteristics that are indicative of hateful or offensive language. The combined feature vector outperformed the existing works. The stylistic features like hate speech terms, capitalization terms and repetitive words played a crucial role in hate speech detection

V. CONCLUSION

In this study we have integrated different feature representation methods to predict hate speech. We have explored TF-IDF, TF-IDF Plus stylistic, word embedding methods Word2Vec and GloVe. We have trained SVM, Random Forest and XGBoost algorithms on those feature representations. Our experimental findings demonstrate that incorporating stylistic features alongside TF-IDF scores yields superior performance compared to existing approaches. Stylistic features such as hate speech terms, capitalization patterns, and repetitive words played a crucial role in the detection of hate speech.

REFERENCES

[1] Taniya Roy, Hindutva’s circulation of anti-muslim hate aided by digital platforms, finds report, 2022.

[2] K. E. Riehm, K. A. Feder, K. N. Tormohlen, R. M. Crum, A. S. Young, K. M. Green, L. R. Pacek, L. N. La Flair, R. Mojtabai, Associations Between Time Spent Using Social Media and Internalizing and Externalizing Problems Among US Youth, JAMA Psychiatry 76 (2019) 1266–1273

[3] T. Ranasinghe, K. Ghosh, A. S. Pal, A. Senapati, A. E. Dmonte, M. Zampieri, S. Modha, S. Satapara, Overview of the HASOC subtracks at FIRE 2023: Hate speech and offensive content identification in assamese, bengali, bodo, gujarati and sinhala, in: Proceedings of the 15th Annual Meeting of the Forum for Information Retrieval Evaluation, FIRE 2023, Goa, India. December 15-18, 2023, ACM, 2023.

[4] R. Guermazi, M. Hammami, A. B. Hamadou, "Using a semi-automatic keyword dictionary for improving violent web site filtering," in Proc. 2007 Third International IEEE Conference on Signal-Image Technologies and Internet-Based System, 2007, pp. 337–344.

[5] N. Djuric, J. Zhou, R. Morris, M. Grbovic, V. Radosavljevic, N. Bhamidipati, "Hate speech detection with comment embeddings," in Proc. 24th International Conference on World Wide Web, WWW ’15 Companion, Association for Computing Machinery, New York, NY, USA, 2015, pp. 29–30.

[6] Adi Narayana Reddy, K., Laskari, N. K., Shyam Chandra Prasad, G., & Sreekanth, N. (2022). Fusion-Based Celebrity Profiling Using Deep Learning. In Intelligent System Design: Proceedings of INDIA 2022 (pp. 107-113). Singapore: Springer Nature Singapore.

[7] P. Badjatiya, S. Gupta, M. Gupta, V. Varma, "Deep learning for hate speech detection in tweets," in Proc. 26th International Conference on World Wide Web Companion, WWW ’17 Companion, International World Wide Web Conferences Steering Committee, Republic and Canton of Geneva, CHE, 2017, pp. 759–760

[8] Kavadi, D. P., Al-Turjman, F., Reddy, K. A. N., & Patan, R. (2021). A machine learning approach for celebrity profiling. International Journal of Ad Hoc and Ubiquitous Computing, 38(1-3), 111-126.

[9] J. A. Leite, D. Silva, K. Bontcheva, C. Scarton, "Toxic language detection in social media for Brazilian Portuguese: New dataset and multilingual analysis," in Proc. 1st Conference of the Asia-Pacific Chapter of the Association for Computational Linguistics and the 10th International Joint Conference on Natural Language Processing, Association for Computational Linguistics, Suzhou, China, 2020, pp. 914–924

[10] A. Saroj, S. Pal, "An Indian language social media collection for hate and offensive speech," in Proc. Workshop on Resources and Techniques for User and Author Profiling in Abusive Language, European Language Resources Association (ELRA), Marseille, France, 2020, pp. 2–8

[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 Proc. 13th International Workshop on Semantic Evaluation, Association for Computational Linguistics, Minneapolis, Minnesota, USA, 2019, pp. 54–63.

[12] A. Ghosh Chowdhury, A. Didolkar, R. Sawhney, R. R. Shah, "ARHNet - leveraging community interaction for detection of religious hate speech in Arabic," in Proc. 57th Annual Meeting of the Association for Computational Linguistics: Student Research Workshop, Association for Computational Linguistics, Florence, Italy, 2019, pp. 273–280.

[13] S. Chopra, R. Sawhney, P. Mathur, R. Ratn Shah, "Hindi-english hate speech detection: Author profiling, debiasing, and practical perspectives," in Proc. AAAI Conference on Artificial Intelligence 34, 2020, pp. 386–393.

[14] R. Kapoor, Y. Kumar, K. Rajput, R. R. Shah, P. Kumaraguru, R. Zimmermann, "Mind your language: Abuse and offense detection for code-switched languages," in Proc. AAAI Conference on Artificial Intelligence 33, 2019, pp. 9951–9952.

[15] A. Vaswani et al., "Attention is all you need," in I. Guyon et al. (Eds.), Advances in Neural Information Processing Systems, vol. 30, Curran Associates, Inc., 2017, pp.

[16] T. Ranasinghe, S. Gupte, M. Zampieri, I. Nwogu, "Wlv-rit at hasoc-dravidian-codemixfire2020: Offensive language identification in code-switched YouTube comments," 2020. arXiv:2011.00559.

[17] I. Kwok, Y. Wang, "Locate the hate: Detecting tweets against blacks," in Proc. Twenty-Seventh AAAI Conference on Artificial Intelligence, AAAI’13, AAAI Press, 2013, pp. 1621–1622.

[18] Y. Chen, Y. Zhou, S. Zhu, H. Xu, "Detecting offensive language in social media to protect adolescent online safety," in Proc. 2012 International Conference on Privacy, Security, Risk and Trust and 2012 International Confernece on Social Computing, 2012, pp. 71–80.

[19] R. Rajalakshmi, Y. Reddy, "Dlrg@hasoc 2020: A hybrid approach for hate and offensive content identification in multilingual tweets," in Proc. Fire, 2020.
[20] J. Pennington, R. Socher, C. Manning, "GloVe: Global vectors for word representation," in Proc. 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP), Association for Computational Linguistics, Doha, Qatar, 2014, pp. 1532–1543.
[21] P. Bojanowski, E. Grave, A. Joulin, T. Mikolov, "Enriching word vectors with subword information," Trans. Assoc. Comput. Linguist., vol. 5, pp. 135-146, 2017
[22] R. Raj, S. Srivastava, S. Saumya, "Nsit & iiitdwd @ hasoc 2020: Deep learning model for hate-speech identification in Indo-European languages," in Fire, 2020.
[23] A. K. Mishra, S. Saumya, A. Kumar, "Iiit_dwd@hasoc 2020: Identifying offensive content in Indo-European languages," in Proc. Work. Notes FIRE 2020 - Forum Inf. Retrieval Eval., Hyderabad, India, December 16-20, 2020, vol. 2826, CEUR Workshop Proceedings, pp. 139-144.
[24] S. Mohtaj, V. Woloszyn, S. Möller, "Tub at hasoc 2020: Character based LSTM for hate speech detection in Indo-European languages," in Fire, 2020.
[25] J. Devlin, M. Chang, K. Lee, K. Toutanova, "BERT: Pre-training of deep bidirectional transformers for language understanding," in Proc. 2019 Conf. North American Chapter Assoc. Comput. Linguist. Hum. Lang. Technol., Minneapolis, MN, USA, June 2-7, 2019, vol. 1, Assoc. Comput. Linguist., pp. 4171-4186. DOI: 10.18653/v1/n19-1423
[26] M. Mozafari, R. Farahbakhsh, N. Crespi, "A BERT-Based Transfer Learning Approach for Hate Speech Detection in Online Social Media," 2019, pp. 928-940..
[27] S. S. Aluru, B. Mathew, P. Saha, A. Mukherjee, "Deep learning models for multilingual hate speech detection," 2020
[28] A. Conneau, K. Khandelwal, N. Goyal, V. Chaudhary, G. Wenzek, F. Guzmán, E. Grave, M. Ott, L. Zettlemoyer, V. Stoyanov, "Unsupervised cross-lingual representation learning at scale," in Proc. 58th Annu. Meeting Assoc. Comput. Linguist., Online, 2020, pp. 8440-8451..
[29] S. Khanuja, D. Bansal, S. Mehtani, S. Khosla, A. Dey, B. Gopalan, D. K. Margam, P. Aggarwal, R. T. Nagipogu, S. Dave, et al., "Muril: Multilingual representations for Indian languages," in arXiv preprint arXiv:2103.10730, 2021.
[30] V. Dhananjaya, P. Demotte, S. Ranathunga, S. Jayasena, "BERTifying Sinhala - A comprehensive analysis of pre-trained language models for Sinhala text classification," in Proc. Thirteenth Lang. Resources Evaluation Conf., Marseille, France, 2022, pp. 7377-7385.
[31] T. Hasan, A. Bhattacharjee, K. Samin, M. Hasan, M. Basak, M. S. Rahman, R. Shahriyar, "Not low-resource anymore: Aligner ensembling, batch filtering, and new datasets for Bengali-English machine translation," in Proc. 2020 Conf. Empirical Methods Nat. Lang. Process., Online, 2020, pp. 2612-2623.
[32] S. Doddapaneni, R. Aralikatte, G. Ramesh, S. Goyal, M. M. Khapra, A. Kunchukuttan, P. Kumar, "Towards leaving no Indic language behind: Building monolingual corpora, benchmark, and models for Indic languages," 2023.
[33] I. M. Moosa, M. E. Akhter, A. B. Habib, "Does transliteration help multilingual language modeling?," in Findings Assoc. Comput. Linguist.: EACL 2023, Dubrovnik, Croatia, 2023, pp. 670-685. DOI: 10.18653/v1/2023.findings-eacl.50.
[34] K. Ghosh, D. A. Senapati, "Hate speech detection: a comparison of mono and multilingual transformer model with cross-language evaluation," in Proc. 36th Pacific Asia Conf. Lang., Inf., Comput., De La Salle University, Manila, Philippines, 2022, pp. 853-865.
[35] S. Satapara, P. Majumder, T. Mandl, S. Modha, H. Madhu, T. Ranasinghe, M. Zampieri, K. North, D. Premasiri, "Overview of the hasoc subtrack at FIRE 2022: Hate speech and offensive content identification in English and Indo-Aryan languages," in Proc. 14th Annu. Meeting Forum Inf. Retrieval Eval., Association for Computing Machinery, New York, NY, USA, 2023, pp. 4-7.
[36] 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 Proc. 13th Annu. Meeting Forum Inf. Retrieval Eval., Association for Computing Machinery, New York, NY, USA, 2022, pp. 1-3..
[37] M. Bhatia, T. S. Bhotia, A. Agarwal, P. Ramesh, S. Gupta, K. Shridhar, F. Laumann, A. Dash, "One to rule them all: Towards joint Indic language hate speech detection," 2021.
[38] T. Ranasinghe, K. North, D. Premasiri, M. Zampieri, "Overview of the hasoc subtrack at FIRE 2022: Offensive language identification in Marathi," 2022.
[39] T. Ranasinghe, I. Anuradha, D. Premasiri, K. Silva, H. Hettiarachchi, L. Uyangodage, M. Zampieri, "Sold: Sinhala offensive language dataset," 2022.
[40] L. Lakshmi, K. Dhana Sree Devi, K. Adi Narayana Reddy, Suresh Kumar Grandhi, Sandeep Kumar Panda. "WOMT: Wasserstein Distribution based minimization of False Positives in Breast Tumor classification using Deep Learning", IEEE Access, 2023.
[41] Proceedings of the 2nd International Conference on Big Data, IoT and Machine Learning", Springer Science and Business Media LLC, 2024.