COMBATANT@TamilNLP-ACL2022: Fine-grained Categorization of Abusive Comments using Logistic Regression

Alamgir Hossain[Ψ], Mahathir Mohammad Bishal[Ψ], Eftekhar Hossain[$],
Omar Sharif[Ψ] and Mohammed Moshiul Hoque[Ψ]

[Ψ]Department of Computer Science and Engineering
[$]Department of Electronics and Telecommunication Engineering
[$][Ψ]Chittagong University of Engineering & Technology, Chattogram-4349, Bangladesh
{u1604069, u1604083}@student.cuet.ac.bd
{omar.sharif, eftekhar.hossain, moshiul_240}@cuet.ac.bd

Abstract

With the widespread usage of social media and effortless internet access, millions of posts and comments are generated every minute. Unfortunately, with this substantial rise, the usage of abusive language has increased significantly in these mediums. This proliferation leads to many hazards such as cyber-bullying, vulgarity, online harassment and abuse. Therefore, it becomes a crucial issue to detect and mitigate the usage of abusive language. This work presents our system developed as part of the shared task to detect the abusive language in Tamil. We employed three machine learning (LR, DT, SVM), two deep learning (CNN+BiLSTM, CNN+BiLSTM with Fast-Text) and a transformer-based model (Indic-BERT). The experimental results show that Logistic regression (LR) and CNN+BiLSTM models outperformed the others. Both Logistic Regression (LR) and CNN+BiLSTM with Fast-Text achieved the weighted F_1-score of 0.39. However, LR obtained a higher recall value (0.44) than CNN+BiLSTM (0.36). This leads us to stand the 2nd rank in the shared task competition.

1 Introduction

With the rapid growth of user-generated content in social media, the emergence of abusive content also increased dramatically (Priyadharshini et al., 2021; Kumaresan et al., 2021). This insurgence has become a reason of worry for governments, policy-makers, social scientists and tech companies since it has detrimental consequences on society (Sharif et al., 2021b; Chakravarthi et al., 2020b). Currently, we are living in an information era where social media plays a vital role in shaping people's minds, and opinions (Perse and Lambe, 2016; Chakravarthi et al., 2021). Therefore mitigating the usage of abusive language has become extremely important (Sharif and Hoque, 2021b). Companies like Facebook, YouTube, Twitter have been trying to achieve

this for years (Ghanghor et al., 2021a,b; Yasaswini et al., 2021). It is impossible to monitor and moderate social media content manually because of its large volume and its messy forms (Meyer, 2016). Therefore, it is necessary to develop an intelligent system to tackle this issue. Several studies have been conducted to detect abusive language for English, and other high resource languages (Kumar et al., 2020; Sampath et al., 2022; Ravikiran et al., 2022; Chakravarthi et al., 2022; Bharathi et al., 2022; Priyadharshini et al., 2022). In contrast, a low-resource language like Tamil remained out of focus and has much room for improvement (Priyadharshini et al., 2020; Chakravarthi et al., 2020a).

Tamil is an official language of Tamil Nadu, Sri Lanka, Singapore, and the Union Territory of Puducherry in India (Anita and Subalalitha, 2019b,a; Subalalitha and Poovammal, 2018). Significant minority speak Tamil in the four other South Indian states of Kerala, Karnataka, Andhra Pradesh, and Telangana, as well as the Union Territory of the Andaman and Nicobar Islands (Subalalitha, 2019; Srinivasan and Subalalitha, 2019; Narasimhan et al., 2018). Tamil, as a Dravidian language, descended from Proto-Dravidian, a proto-language, according to Bhadriraju Krishnamurti. Linguistic reconstruction implies that Proto-Dravidian was spoken about the third millennium BC, likely in the peninsular Indian region surrounding the lower Godavari river basin. The material evidence implies that the speakers of Proto-Dravidian belonged to the civilization linked with South India's Neolithic complexes. The earliest Old Tamil documents are small inscriptions in Adichanallur dating from 905 BC to 696 BC. Tamil has the most ancient non-Sanskritic Indian literature (Sakuntharaj and Mahesan, 2021, 2017, 2016; Thavareesan and Mahesan, 2019, 2020a,b, 2021).

This work aims to build a system that can classify abusive language from Tamil text concerning

eight different categories: *Hope-Speech, Homophobia, Misandry, Counter-speech, Misogyny, Xenophobia, Transphobia and None-of-the-above*. Various machine learning (ML), deep learning (DL), and transformer-based models have been used to attain this goal. The key contributions of this work are illustrated in the following:

- Developed multiple ML and DL techniques to classify abusive texts in Tamil into eight classes.

- Investigated the performance of the models to find the suitable method for the classification of abusive comments and analyzed in-depth error, providing useful insight into abusive text classification.

2 Related Work

Recently, researchers are trying to develop methods and tools to analyze social media sites like Twitter, Facebook, and Snapchat since these mediums has become integral part of our life (Anand and Eswari, 2019). Studies have already been conducted to detect abusive or offensive comments on social media (Sharif et al., 2021a; Aurpa et al., 2022; Sharif et al., 2020). Few researches has focused on other overlapping domains such as hate speech (Founta et al., 2018; Waseem et al., 2017), cyberbullying (Fosler-Lussier et al., 2012), racism/sexism (Talat et al., 2018), aggression & trolling (Zampieri et al., 2019) and so on. All of these researches primarily conducted for high-resource languages. Very few researches have been carried out to detect abusive language for low-resources languages like Tamil. Eshan and Hasan (2017) evaluated the effectiveness of RF, NB, and SVM classifiers to detect abusive language. Their system achieved the maximum accuracy ($\approx 95\%$) for SVM with linear kernel and tri-gram features. Ishmam and Sharmin (2019) collected roughly 5000 Bengali abusive comments from Facebook and categorized them into six different classes: *hate speech, communal attack, inciteful comments, religious hatred, political hatred etc*. They obtained the highest accuracy of 70.10% utilizing the GRU-based model. Salminen et al. (2020) collected 197,566 comments from Twitter, Wikipedia, Reedit and YouTube, where 20% of the data was hateful. They applied logistic regression, naïve bayes, support vector machines, XGBoost techniques on this dataset and obtained 0.92 F_1-score using XGBoost classi-

fier. Sharif and Hoque (2021a) developed a gold standard dataset on Bengali aggressive comments from social media called 'ATxtC', which contains 7591 annotated data. In the subsequent work they presented a novel Bengali aggressive text dataset (called 'BAD') with two-level of annotation (Sharif and Hoque, 2021b). They proposed a weighted ensemble technique that uses m-BERT, distil-BERT, Bangla-BERT, and XLM-R as base classifiers to identify and categorize aggressive texts in Bengal. The model achieved the highest weighted-score of 93.43% in the identification task and 93.11% in the categorization task.

3 Task and Dataset Description

Task organizers created a gold standard dataset to detect abusive comments from Tamil social networking sites. This task aims to develop a system that can correctly identify abusive texts from a given set of texts in Tamil. We used the corpus provided by the organizers of the workshop[1] (Chakravarthi, 2020; Chakravarthi and Muralidaran, 2021; Hande et al., 2021; Priyadharshini et al., 2022). The shared task required classifying a text into eight predefined classes (i.e., Misogyny, Misandry, Homophobia, Transphobia, Xenophobia, Counter-speech, Hope-speech and None-of-the-above). Table 1 reports the number of samples in the train, validation, and test sets for each class. Dataset is quite imbalanced where Transphobia and Homophobia classes have only 10 and 51 text samples, respectively. Before model development, we preprocessed the dataset to exclude irrelevant characters, numbers, symbols, punctuation marks, and emojis.

Class	Train	Validation	Test
Misogyny	125	24	48
Misandry	447	104	127
Homophobia	35	8	8
Transphobia	6	2	2
Xenophobia	95	29	25
Counter-speech	149	36	47
Hope-speech	87	11	26
None-of-the-above	1296	346	416
Total	2240	560	699

Table 1: Class wise dataset distribution in train, validation and test set

---

[1]https://competitions.codalab.org/competitions/36403

4 Methodology

The techniques and methods used to detect abusive categories for given YouTube comments are briefly explained in this section. We cleaned the raw data first by stripping away noisy elements and then extracted features (Lewis, 1992) using various feature extraction techniques, including TF-IDF, Word2Vec, and FastText. We used ML and DL based techniques for the baseline evaluation. The schematic process of our approach is depicted in Figure 1.
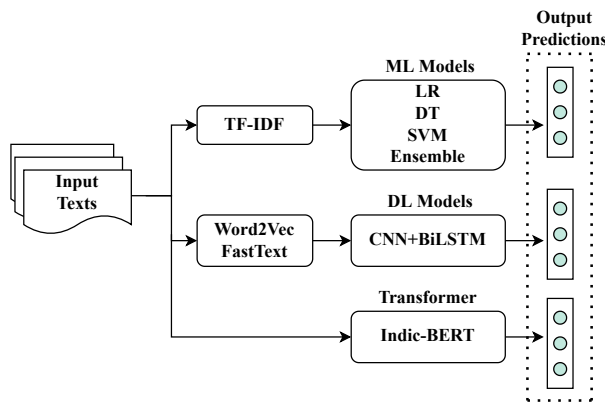

Figure 1: Schematic process of abusive comments classification

4.1 Feature Extraction

Feature extraction is conducted prior to training the models. The TF-IDF (Nayel, 2020) values of the unigram features are calculated and used for training ML models. On the other hand, Word2Vec (Jurgens, 2021) and FastText (Joulin et al., 2017) embeddings are used as feature for the DL models. Keras embedding layer generates the embedding vectors of the dimension of 100. In contrast, a pre-trained embedding matrix is in the case of FastText embedding.

4.2 ML Baselines

In order to design the abusive comment detection system, we developed several ML-based methods such as logistic regression (LR), decision tree (DT), and support vector machine (SVM). After constructing the three models, we also use the majority voting ensemble technique to predict the abusive category of the texts. Furthermore, in search of improved performance, an ensemble approach is applied using the classifiers mentioned earlier. In LR and DT models, the C value is settled at 2, whereas SVM is implemented with a C value of 6.

4.3 DL Baselines

In the case of DL approach (Ruiz et al., 2020), we combined CNN and LSTM (Du et al., 2020) to classify a given comment. A total of seven layers is used to construct the combined model. Initially, a sequence vector of length 260 is fed to the embedding layer. Subsequently, two convolution layers are added with the 'relu' activation function. Features are downsampled through a max-pooling layer before passing to the BiLSTM layer. BiLSTM has 128 units, and the overfitting problem is reduced by setting the dropout rate to 0.2. Finally, a softmax layer is used to get the predictions. We also performed experimentation with pre-trained word vectors (FastText). We use the 'Adam' optimizer with a learning rate of $1e^{-3}$, and a loss function of 'sparse_categorical_crossentropy'. The model has been trained for 25 epochs with a batch size of 32.

4.3.1 Transformers

Considering the recent vogue of transformers, we also employ a transformer-based model. Specifically, we chose Indic-BERT as a pre-trained model is trained on the texts of various Indian languages such as Tamil, Bangla, and Telugu. We chose Indic-BERT because it has far fewer parameters than other multilingual models (i.e., mBERT, XLM-R, etc.) while achieving comparable performance (Kakwani et al., 2020). The maximum length of the input text is settled to 150 and use Ktrain (Maiya, 2020) package to fine-tune the model. The model is compiled using the Ktrain 'fit_onecycle' method along with a learning rate of $2e^{-5}$. Finally, the training is performed for 4 epochs bypassing 12 instances at each iteration. The implementation details of implemented models have been open sourced for reproducibility[2].

5 Results and Analysis

The performance of the various methods on the test set is reported in Table 2. The macro F_1-score measures the supremacy of the models. However, we pay close attention to the other measures such as accuracy (A), precision (P) and recall (R) scores.

It is observed that, among the ML models, the LR model outperformed the DT and SVM models achieving the highest macro F_1-score (0.39). The combination of CNN and BiLSTM (C+B) achieved a very low macro F_1-score (0.16) when trained with

---

[2]https://github.com/m1n1-coder/ML-and-DL-models-of-Tamil-Abusive-Comment-Detection

Methods	Classifier	P	R	F_1-score	A
ML models	LR	0.38	**0.44**	**0.39**	0.60
	DT	0.31	0.34	0.32	0.57
	SVM	0.54	0.26	0.29	0.66
	Ensemble	0.26	0.44	0.28	0.67
DL models	C+B (Word2Vec)	0.19	0.18	0.16	0.31
	C+B (FastText)	0.52	0.36	0.39	0.63
Transformer	Indic-BERT	0.22	0.20	0.19	0.69

Table 2: Performance of various models on the test set. Here, C+B represents the CNN+BiLSTM model

Word2Vec features. Surprisingly, the performance is improved to 0.39 when we used a pre-trained word embedding (i.e. FastText). Unfortunately, the transformer model, Indic-BERT, could not provide satisfactory performance on the test set. Moreover, we conducted a thorough investigation into all of the employed models. The outcomes of the investigation is presented in Table 3. It is revealed that the LR model predicts 6 of the 8 categories with the highest F_1-score. This demonstrated that the LR model performed admirably and was the best model across all evaluation metrics. However, CNN+BiLSTM with FastText embedding also achieved the same macro F_1-score (0.39). On the other hand, the transformer model performs poorly due to the prevalence of local words across the different abusive classes.

The comparative analysis illustrates that our model (i.e., **COMBATANT**) achieved the 2nd position in the task (Table-4). Although we investigated various ML and DL models on the corpus, the submission included the best three models (LR, SVM, and CNN+BiLSTM (with FastText)). The LR model outperformed the other models by achieving the highest F_1-score.

5.1 Error Analysis

The LR classifier outperformed all models in classifying Tamil abusive comments on the shared task dataset. However, it is necessary to investigate the errors of the model in order to assess how accurately the classifier performed across the different classes. The confusion matrix is used to illustrate the errors (Figure 2). We noticed that, among the classes, **Misandry** and **Counter-Speech** contained a relatively high true positive rate (TPR). Misandry class obtained a TPR of 65.35%, whereas Counter-Speech achieved 53.2%. However, **Transphobia** has a TPR of 0%. With a low TPR, **Homophobia** class also experienced a large number of miss-
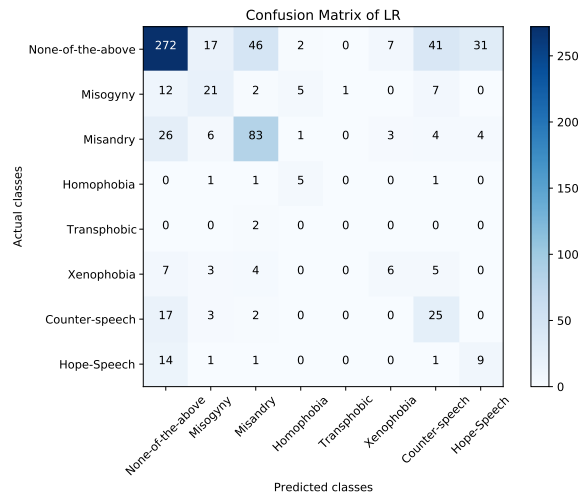

Figure 2: Confusion matrix of the best model (LR).

classification. This lower outcome could be occurred due to inadequacy and class imbalance of data. As a result, many of the test data were incorrectly classified as **None-of-the-above**.

6   Conclusion

This paper presents the various models developed to classify abusive comments in Tamil. This work used three ML, two DL classifiers and one transformer-based model to perform the classification task. The LR model with TF-IDF features outperformed all models by obtaining the highest macro F_1-score (0.39). Although the combined CNN and BiLSTM model (C+B) achieved a similar macro F_1-score (0.39) with FastText features, the LR model obtained a higher recall value (0.44). Surprisingly, Indic-BERT performed poorly compared to the ML and DL models. These inferior results might occur because of the prevalence of local words, which is unknown to the model. It will be interesting to investigate how the model performs if the dataset is used in more advanced transformer models (XML-R, Electra, mBERT, MuRIL).

Classes	LR	DT	SVM	Ensemble	C+B(Word2Vec)	C+B(FastText)	Indic-BERT
Misogyny	**0.42**	0.31	0.21	0.24	0.08	0.14	0.15
Misandry	**0.62**	0.50	0.54	0.55	0.28	0.53	0.42
Homophobia	**0.48**	0.42	0.46	0.43	0.00	0.50	0.15
Transphobic	0.00	0.00	0.00	0.00	0.00	**0.67**	0.03
Xenophobia	**0.29**	0.14	0.07	0.07	0.10	0.10	0.07
Counter-speech	**0.38**	0.28	0.15	0.00	0.21	0.26	0.11
Hope-Speech	**0.26**	0.18	0.14	0.20	0.11	0.14	0.10
None-of-the-above	0.71	0.72	0.78	**0.79**	0.47	0.78	0.49

Table 3: Class-wise performance of models in terms of F_1-score

Team_Names	Precision	Recall	F1-score	Rank
CEN-Tamil	0.380	0.290	0.320	1
COMBATANT	0.290	0.330	0.300	2
DE-ABUSE	0.330	0.290	0.291	3
DLRG	0.340	0.260	0.270	4
TROOPER	0.400	0.230	0.250	5
abusive-checker	0.140	0.140	0.140	6
Optimize_Prime_Tamil_Run1	0.130	0.130	0.130	7
GJG_Tamil	0.130	0.140	0.130	8
umuteam_tamil	0.130	0.130	0.130	9
MUCIC	0.120	0.130	0.120	10
BpHigh_tamil(1)	0.180	0.120	0.060	11
SSNCSE_NLP	0.130	0.140	0.090	12

Table 4: Summary of performance comparison for all participating teams in the shared task

Furthermore, we aim to tackle the data imbalance problem by adding more diverse data to the existing corpus that might improve the model's performance.

Acknowledgements

References

Mukul Anand and R Eswari. 2019. Classification of abusive comments in social media using deep learning. In *2019 3rd international conference on computing methodologies and communication (ICCMC)*, pages 974–977. IEEE.

R Anita and CN Subalalitha. 2019a. An approach to cluster Tamil literatures using discourse connectives. In *2019 IEEE 1st International Conference on Energy, Systems and Information Processing (ICESIP)*, pages 1–4. IEEE.

R Anita and CN Subalalitha. 2019b. Building discourse parser for Thirukkural. In *Proceedings of the 16th International Conference on Natural Language Processing*, pages 18–25.

Tanjim Taharat Aurpa, Rifat Sadik, and Md Shoaib Ahmed. 2022. Abusive bangla comments detection on facebook using transformer-based deep learning models. *Social Network Analysis and Mining*, 12(1):1–14.

B Bharathi, Bharathi Raja Chakravarthi, Subalalitha Chinnaudayar Navaneethakrishnan, N Sripriya, Arunaggiri Pandian, and Swetha Valli. 2022. Findings of the shared task on Speech Recognition for Vulnerable Individuals in Tamil. In *Proceedings of the Second Workshop on Language Technology for Equality, Diversity and Inclusion*. Association for Computational Linguistics.

Bharathi Raja Chakravarthi. 2020. Hopeedi: A multilingual hope speech detection dataset for equality, diversity, and inclusion. In *Proceedings of the Third Workshop on Computational Modeling of People's Opinions, Personality, and Emotion's in Social Media*, pages 41–53.

Bharathi Raja Chakravarthi and Vigneshwaran Muralidaran. 2021. Findings of the shared task on hope speech detection for equality, diversity, and inclusion. In *Proceedings of the First Workshop on Language Technology for Equality, Diversity and Inclusion*, pages 61–72, Kyiv. Association for Computational Linguistics.

Bharathi Raja Chakravarthi, Vigneshwaran Muralidaran, Ruba Priyadharshini, and John Philip McCrae. 2020a. Corpus creation for sentiment analysis in code-mixed Tamil-English text. In *Proceedings of the 1st Joint Workshop on Spoken Language Technologies for Under-resourced languages (SLTU) and Collaboration and Computing for Under-Resourced Languages (CCURL)*, pages 202–210, Marseille, France. European Language Resources association.

Bharathi Raja Chakravarthi, Ruba Priyadharshini, Thenmozhi Durairaj, John Phillip McCrae, Paul Buitaleer, Prasanna Kumar Kumaresan, and Rahul Ponnusamy.

2022. Findings of the shared task on Homophobia Transphobia Detection in Social Media Comments. In *Proceedings of the Second Workshop on Language Technology for Equality, Diversity and Inclusion*. Association for Computational Linguistics.

Bharathi Raja Chakravarthi, Ruba Priyadharshini, Vigneshwaran Muralidaran, Shardul Suryawanshi, Navya Jose, Elizabeth Sherly, and John P McCrae. 2020b. Overview of the track on sentiment analysis for Dravidian languages in code-mixed text. In *Forum for Information Retrieval Evaluation*, pages 21–24.

Bharathi Raja Chakravarthi, Ruba Priyadharshini, Rahul Ponnusamy, Prasanna Kumar Kumaresan, Kayalvizhi Sampath, Durairaj Thenmozhi, Sathiyaraj Thangasamy, Rajendran Nallathambi, and John Phillip McCrae. 2021. Dataset for identification of homophobia and transophobia in multilingual YouTube comments. *arXiv preprint arXiv:2109.00227*.

Jinhua Du, Yan Huang, and Karo Moilanen. 2020. Pointing to select: A fast pointer-LSTM for long text classification. In *Proceedings of the 28th International Conference on Computational Linguistics*, pages 6184–6193, Barcelona, Spain (Online). International Committee on Computational Linguistics.

Shahnoor C Eshan and Mohammad S Hasan. 2017. An application of machine learning to detect abusive bengali text. In *2017 20th International Conference of Computer and Information Technology (ICCIT)*, pages 1–6. IEEE.

Eric Fosler-Lussier, Ellen Riloff, and Srinivas Bangalore. 2012. Proceedings of the 2012 conference of the north american chapter of the association for computational linguistics: Human language technologies. In *Proceedings of the 2012 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*.

Antigoni Maria Founta, Constantinos Djouvas, Despoina Chatzakou, Ilias Leontiadis, Jeremy Blackburn, Gianluca Stringhini, Athena Vakali, Michael Sirivianos, and Nicolas Kourtellis. 2018. Large scale crowdsourcing and characterization of twitter abusive behavior. In *Twelfth International AAAI Conference on Web and Social Media*.

Nikhil Ghanghor, Parameswari Krishnamurthy, Sajeetha Thavareesan, Ruba Priyadharshini, and Bharathi Raja Chakravarthi. 2021a. IIITK@DravidianLangTech-EACL2021: Offensive language identification and meme classification in Tamil, Malayalam and Kannada. In *Proceedings of the First Workshop on Speech and Language Technologies for Dravidian Languages*, pages 222–229, Kyiv. Association for Computational Linguistics.

Nikhil Ghanghor, Rahul Ponnusamy, Prasanna Kumar Kumaresan, Ruba Priyadharshini, Sajeetha Thavareesan, and Bharathi Raja Chakravarthi. 2021b.

IIITK@LT-EDI-EACL2021: Hope speech detection for equality, diversity, and inclusion in Tamil , Malayalam and English. In *Proceedings of the First Workshop on Language Technology for Equality, Diversity and Inclusion*, pages 197–203, Kyiv. Association for Computational Linguistics.

Adeep Hande, Ruba Priyadharshini, Anbukkarasi Sampath, Kingston Pal Thamburaj, Prabakaran Chandran, and Bharathi Raja Chakravarthi. 2021. Hope speech detection in under-resourced kannada language.

Alvi Md Ishmam and Sadia Sharmin. 2019. Hateful speech detection in public facebook pages for the bengali language. In *2019 18th IEEE International Conference On Machine Learning And Applications (ICMLA)*, pages 555–560. IEEE.

Armand Joulin, Edouard Grave, Piotr Bojanowski, and Tomas Mikolov. 2017. Bag of tricks for efficient text classification. In *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 2, Short Papers*, pages 427–431, Valencia, Spain. Association for Computational Linguistics.

David Jurgens. 2021. Learning about word vector representations and deep learning through implementing word2vec. In *Proceedings of the Fifth Workshop on Teaching NLP*, pages 108–111, Online. Association for Computational Linguistics.

Divyanshu Kakwani, Anoop Kunchukuttan, Satish Golla, Gokul N.C., Avik Bhattacharyya, Mitesh M. Khapra, and Pratyush Kumar. 2020. IndicNLPSuite: Monolingual Corpora, Evaluation Benchmarks and Pre-trained Multilingual Language Models for Indian Languages. In *Findings of EMNLP*.

Ritesh Kumar, Atul Kr. Ojha, Bornini Lahiri, Marcos Zampieri, Shervin Malmasi, Vanessa Murdock, and Daniel Kadar, editors. 2020. *Proceedings of the Second Workshop on Trolling, Aggression and Cyberbullying*. European Language Resources Association (ELRA), Marseille, France.

Prasanna Kumar Kumaresan, Ratnasingam Sakuntharaj, Sajeetha Thavareesan, Subalalitha Navaneethakrishnan, Anand Kumar Madasamy, Bharathi Raja Chakravarthi, and John P McCrae. 2021. Findings of shared task on offensive language identification in Tamil and Malayalam. In *Forum for Information Retrieval Evaluation*, pages 16–18.

David D. Lewis. 1992. Feature selection and feature extraction for text categorization. In *Speech and Natural Language: Proceedings of a Workshop Held at Harriman, New York, February 23-26, 1992*.

Arun S Maiya. 2020. ktrain: A low-code library for augmented machine learning. *arXiv preprint arXiv:2004.10703*.

Robinson Meyer. 2016. Twitter's famous racist problem. *The Atlantic*, 7:21.

Anitha Narasimhan, Aarthy Anandan, Madhan Karky, and CN Subalalitha. 2018. Porul: Option generation and selection and scoring algorithms for a tamil flash card game. *International Journal of Cognitive and Language Sciences*, 12(2):225–228.

Hamada Nayel. 2020. NAYEL at SemEval-2020 task 12: TF/IDF-based approach for automatic offensive language detection in Arabic tweets. In *Proceedings of the Fourteenth Workshop on Semantic Evaluation*, pages 2086–2089, Barcelona (online). International Committee for Computational Linguistics.

Elizabeth M Perse and Jennifer Lambe. 2016. *Media effects and society*. Routledge.

Ruba Priyadharshini, Bharathi Raja Chakravarthi, Subalalitha Chinnaudayar Navaneethakrishnan, Thenmozhi Durairaj, Malliga Subramanian, Kogilavani Shanmugavadivel, Siddhanth U Hegde, and Prasanna Kumar Kumaresan. 2022. Findings of the shared task on Abusive Comment Detection in Tamil. In *Proceedings of the Second Workshop on Speech and Language Technologies for Dravidian Languages*. Association for Computational Linguistics.

Ruba Priyadharshini, Bharathi Raja Chakravarthi, Sajeetha Thavareesan, Dhivya Chinnappa, Durairaj Thenmozhi, and Rahul Ponnusamy. 2021. Overview of the DravidianCodeMix 2021 shared task on sentiment detection in Tamil, Malayalam, and Kannada. In *Forum for Information Retrieval Evaluation*, pages 4–6.

Ruba Priyadharshini, Bharathi Raja Chakravarthi, Mani Vegupatti, and John P McCrae. 2020. Named entity recognition for code-mixed Indian corpus using meta embedding. In *2020 6th international conference on advanced computing and communication systems (ICACCS)*, pages 68–72. IEEE.

Manikandan Ravikiran, Bharathi Raja Chakravarthi, Anand Kumar Madasamy, Sangeetha Sivanesan, Ratnavel Rajalakshmi, Sajeetha Thavareesan, Rahul Ponnusamy, and Shankar Mahadevan. 2022. Findings of the shared task on Offensive Span Identification in code-mixed Tamil-English comments. In *Proceedings of the Second Workshop on Speech and Language Technologies for Dravidian Languages*. Association for Computational Linguistics.

Guillermo Ruiz, Eric S. Tellez, Daniela Moctezuma, Sabino Miranda-Jiménez, Tania Ramírez-delReal, and Mario Graff. 2020. Infotec + CentroGEO at SemEval-2020 task 8: Deep learning and text categorization approach for memes classification. In *Proceedings of the Fourteenth Workshop on Semantic Evaluation*, pages 1141–1147, Barcelona (online). International Committee for Computational Linguistics.

Ratnasingam Sakuntharaj and Sinnathamby Mahesan. 2016. A novel hybrid approach to detect and correct spelling in Tamil text. In *2016 IEEE International Conference on Information and Automation for Sustainability (ICIAfS)*, pages 1–6.

Ratnasingam Sakuntharaj and Sinnathamby Mahesan. 2017. Use of a novel hash-table for speeding-up suggestions for misspelt Tamil words. In *2017 IEEE International Conference on Industrial and Information Systems (ICIIS)*, pages 1–5.

Ratnasingam Sakuntharaj and Sinnathamby Mahesan. 2021. Missing word detection and correction based on context of Tamil sentences using n-grams. In *2021 10th International Conference on Information and Automation for Sustainability (ICIAfS)*, pages 42–47.

Joni Salminen, Maximilian Hopf, Shammur A Chowdhury, Soon-gyo Jung, Hind Almerekhi, and Bernard J Jansen. 2020. Developing an online hate classifier for multiple social media platforms. *Human-centric Computing and Information Sciences*, 10(1):1–34.

Anbukkarasi Sampath, Thenmozhi Durairaj, Bharathi Raja Chakravarthi, Ruba Priyadharshini, Subalalitha Chinnaudayar Navaneethakrishnan, Kogilavani Shanmugavadivel, Sajeetha Thavareesan, Sathiyaraj Thangasamy, Parameswari Krishnamurthy, Adeep Hande, Sean Benhur, Kishor Kumar Ponnusamy, and Santhiya Pandiyan. 2022. Findings of the shared task on Emotion Analysis in Tamil. In *Proceedings of the Second Workshop on Speech and Language Technologies for Dravidian Languages*. Association for Computational Linguistics.

Omar Sharif and Mohammed Moshiul Hoque. 2021a. Identification and classification of textual aggression in social media: resource creation and evaluation. In *International Workshop on Combating On line Ho st ile Posts in Regional Languages dur ing Emerge ncy Si tuation*, pages 9–20. Springer.

Omar Sharif and Mohammed Moshiul Hoque. 2021b. Tackling cyber-aggression: Identification and fine-grained categorization of aggressive texts on social media using weighted ensemble of transformers. *Neurocomputing*.

Omar Sharif, Mohammed Moshiul Hoque, A. S. M. Kayes, Raza Nowrozy, and Iqbal H. Sarker. 2020. Detecting suspicious texts using machine learning techniques. *Applied Sciences*, 10(18).

Omar Sharif, Eftekhar Hossain, and Mohammed Moshiul Hoque. 2021a. Combating hostility: Covid-19 fake news and hostile post detection in social media.

Omar Sharif, Eftekhar Hossain, and Mohammed Moshiul Hoque. 2021b. NLP-CUET@DravidianLangTech-EACL2021: Offensive language detection from multilingual code-mixed text using transformers. pages 255–261.

R Srinivasan and CN Subalalitha. 2019. Automated named entity recognition from tamil documents. In *2019 IEEE 1st International Conference on Energy,*

Systems and Information Processing (ICESIP), pages 1–5. IEEE.

C. N. Subalalitha. 2019. Information extraction framework for Kurunthogai. *Sādhanā*, 44(7):156.

CN Subalalitha and E Poovammal. 2018. Automatic bilingual dictionary construction for Tirukural. *Applied Artificial Intelligence*, 32(6):558–567.

Zeerak Talat, James Thorne, and Joachim Bingel. 2018. Bridging the gaps: Multi task learning for domain transfer of hate speech detection. In *Online harassment*, pages 29–55. Springer.

Sajeetha Thavareesan and Sinnathamby Mahesan. 2019. Sentiment analysis in Tamil texts: A study on machine learning techniques and feature representation. In *2019 14th Conference on Industrial and Information Systems (ICIIS)*, pages 320–325.

Sajeetha Thavareesan and Sinnathamby Mahesan. 2020a. Sentiment lexicon expansion using Word2vec and fastText for sentiment prediction in Tamil texts. In *2020 Moratuwa Engineering Research Conference (MERCon)*, pages 272–276.

Sajeetha Thavareesan and Sinnathamby Mahesan. 2020b. Word embedding-based part of speech tagging in Tamil texts. In *2020 IEEE 15th International Conference on Industrial and Information Systems (ICIIS)*, pages 478–482.

Sajeetha Thavareesan and Sinnathamby Mahesan. 2021. Sentiment analysis in Tamil texts using k-means and k-nearest neighbour. In *2021 10th International Conference on Information and Automation for Sustainability (ICIAfS)*, pages 48–53.

Zeerak Waseem, Thomas Davidson, Dana Warmsley, and Ingmar Weber. 2017. Understanding abuse: A typology of abusive language detection subtasks. *arXiv preprint arXiv:1705.09899*.

Konthala Yasaswini, Karthik Puranik, Adeep Hande, Ruba Priyadharshini, Sajeetha Thavareesan, and Bharathi Raja Chakravarthi. 2021. IIITT@DravidianLangTech-EACL2021: Transfer learning for offensive language detection in Dravidian languages. In *Proceedings of the First Workshop on Speech and Language Technologies for Dravidian Languages*, pages 187–194, Kyiv. Association for Computational Linguistics.

Marcos Zampieri, Shervin Malmasi, Preslav Nakov, Sara Rosenthal, Noura Farra, and Ritesh Kumar. 2019. Semeval-2019 task 6: Identifying and categorizing offensive language in social media (offenseval). *arXiv preprint arXiv:1903.08983*.