Findings of the Shared Task on Multimodal Sentiment Analysis and Troll Meme Classification in Dravidian Languages

Premjith B[1], Bharathi Raja Chakravarthi[2], Malliga Subramanian[3], Bharathi B[4], Soman KP[1],
Dhanalakshmi Vadivel[4], Sreelakshmi K[1], Arunaggiri Pandian[6], and Prasanna Kumar Kumaresan[7]

[1]Center for Computational Engineering and Networking (CEN),
Amrita School of Engineering, Coimbatore, Amrita Vishwa Vidyapeetham, India
[2]National University of Ireland Galway, [3]Kongu Engineering College, Tamil Nadu, India
[4]SSN College of Engineering, Tamil Nadu, India, [5]RV Government Arts college , Chengalpattu
[6]Thiagarajar College of Engineerin, Madurai, India
[7]Indian Institute of Information Technology and Management, Kerala
`b_premjith@cb.amrita.edu, bharathi.raja@insight-centre.org`

Abstract

This paper presents the findings of the shared task on Multimodal Sentiment Analysis and Troll meme classification in Dravidian languages held at ACL 2022. Multimodal sentiment analysis deals with the identification of sentiment from video. In addition to video data, the task requires the analysis of corresponding text and audio features for the classification of movie reviews into five classes. We created a dataset for this task in Malayalam and Tamil. The Troll meme classification task aims to classify multimodal Troll memes into two categories. This task assumes the analysis of both text and image features for making better predictions. The performance of the participating teams was analysed using the F1-score. Only one team submitted their results in the Multimodal Sentiment Analysis task, whereas we received six submissions in the Troll meme classification task. The only team that participated in the Multimodal Sentiment Analysis shared task obtained an F1-score of 0.24. In the Troll meme classification task, the winning team achieved an F1-score of 0.596.

1 Introduction

People use different modes of content, including video, audio and text, to express their opinion or attitude or interact with another person. These types of data are too complex to process because of the ambiguity at various levels. Moreover, such types data are more user-centric and contextual (Schreck and Keim, 2012). The complexity of the computational processing of social media is more for multimodal data, which includes video, audio and text modalities. The machine learning or deep learning model should utilize/extract the features identified from all modalities for better

decision making (Chakravarthi, 2020; Kumaresan et al., 2021; Chakravarthi and Muralidaran, 2021; Ravikiran et al., 2022; Chakravarthi et al., 2022; Bharathi et al., 2022; Priyadharshini et al., 2022). This shared task on multimedia social media analysis in Dravidian languages includes two subtasks - Multimodal Sentiment Analysis and Troll meme classification.

Sentiment analysis is a Natural Language Processing (NLP) task to identify the underlying sentiment or opinion about movies, products, government policies etc., expressed by people through various social media platforms (Priyadharshini et al., 2021), (Chakravarthi et al., 2021b). Nowadays, social media users use multiple modalities such as images containing text and videos to express their opinions. It has become common to share review videos about movies and products on YouTube, and Facebook (Castro et al., 2019), (Qiyang and Jung, 2019). Therefore, analysis of multiple modalities has become significant in identifying sentiments from video data. Video data contain modalities such as video frames, speech signals, and text (transcripts). Training a machine learning model requires considering the features of the above three modalities to train a machine learning model.

The shared task on multimodal sentiment analysis aims at inviting researchers for developing machine learning models to draw out the sentiments from movie review videos in Malayalam and Tamil. Malayalam and Tamil are members of the Dravidian language family (Sakuntharaj and Mahesan, 2021, 2017, 2016; Thavareesan and Mahesan, 2019, 2020a,b, 2021). In addition to that, both languages are morphologically rich and agglutinative (Premjith and Soman, 2021), (Premjith et al., 2019). This makes these languages complex for

computational processing (Anita and Subalalitha, 2019b,a; Subalalitha and Poovammal, 2018; Subalalitha, 2019; Srinivasan and Subalalitha, 2019; Narasimhan et al., 2018). Moreover, the complexity increases when we deal with speech and text for learning feature representation from video data. We created the dataset (Chakravarthi et al., 2021a) by downloading movie review videos that have been uploaded on YouTube. The Malayalam dataset consisted of 70 videos, and the Tamil dataset contained 64 videos in it. Videos are classified into five categories: Highly Positive, Positive, Neutral, Negative and Highly Negative by considering the facial expressions of the reviewers and the words used to give a review for the movies.

Memes have become prevalent in social media in recent times. People use memes, which can come in image and video modalities, to express their opinions or attitude on various issues (Suryawanshi and Chakravarthi, 2021a). Some memes are created only for entertainment purposes. However, certain users use memes to inflame individuals or organizations. The computational processing of the meme is challenging because the text that appears in an image or video is difficult to extract (Ghanghor et al., 2021a,b; Yasaswini et al., 2021). In addition to that, the text may appear in different fonts and sizes, increasing the complexity. The objective of the Troll meme classification shared task is to classify a Troll meme into either troll or not-troll categories. We collected Troll memes from social media platforms, namely Facebook, Instagram, Whatsapp, and Pinterest (Suryawanshi et al., 2020a).

This paper presents an overview of the shared task on multimodal sentiment analysis and Troll meme classification in Dravidian languages. We considered data collected from both Malayalam and Tamil for these two tasks. This work also discusses the results of the participating teams. We shared the training and validation data with labels and test data without labels with the participants. The participants submitted the predicted labels for the test data by building machine learning models for the task. In total, 56 teams registered for the multimodal sentiment analysis competition organized by Codalab [1]. However, only one team submitted their results for Tamil test data, whereas six teams participated in the Troll meme classification task.

The paper is organized into four sections, including the introduction. Section 2 describes the Multimodal Sentiment Analysis shared task and the dataset used in detail. It is followed by Section 3, which summarizes the Troll meme classification task. In Section 4, we discuss the systems and methodologies submitted by participants for both shared tasks and analysis of the results of submitted models. A summary of the tasks and future work are presented in Section 5.

2 Multimodal Sentiment Analysis Task Description

Artificial intelligence models that can recognize emotion and opinions are helpful for a variety of industries to understand user requirements, preferences and reviews. From virtual assistants to content moderation, sentiment analysis has many applications. A great deal of multimodal content has been published in local languages on social media about products making the task of sentiment analysis challenging (Chakravarthi et al., 2021a). The reach of visual information and the traction of smartphones have facilitated people to use videos for sharing their opinions. Hence, our task aims to identify Highly positive, Positive, Neutral, Negative, and Highly Negative videos by taking the video, speech and text modalities.

2.1 Dataset Description

The dataset consists of videos collected from YouTube and annotated manually. The dataset has 70 Malayalam, and 64 Tamil videos split into training, validation and test set. Each video segment includes manual transcription aligned along with sentiment annotation by volunteer annotators. Videos are annotated into five labels.

- **Highly positive state**: Video clip where the reviewer uses overstated words or expressions

- **Positive**: A video clip where reviewer uses positive words with mild facial expressions to give reviews .

- **Neutral**: There is no explicit or implicit indicator of the speaker's emotional state. Examples are asking for like or subscription or questions about the release date or movie dialogue.

- **Negative**: Videos where negative words and

Class Label	Malayalam	Tamil
Highly Positive	9	8
Positive	39	38
Neutral	8	8
Negative	12	5
Highly Negative	2	5
Total	70	64

Table 1: Number of videos in each class, and language.

Dataset	Malayalam	Tamil
Train	50	44
Validation	10	10
Test	10	10

Table 2: Distribution of Multimodal Sentiment Analysis data in Malayalam and Tamil.

sarcastic comments with soft facial expressions are used.

- **Highly Negative**: Videos where exaggerated negative words with a sullen face and taut voice, are used.

Table 1 gives the various classes and statistics of data points in each class, and Table 2 shows the number of data points in train, test and validation datasets used in both languages.

3 Troll meme Classification Task Description

This shared task aims to devise methodologies and models for detecting troll memes in Tamil using textual and visual features. A Troll meme consists of offensive text and non-offensive images and vice versa, intending to degrade, provoke, or offend a person or a group. A dataset with two classes, namely Troll and Not Troll, has been provided for this shared task. The dataset has images associated with captions and comprises memes and transcribed text in Latin, which are annotated and transcribed. The participants used images, captions, or both to do the classification. A sample troll and non-troll meme has been shown in Figure. The task organisers have enhanced the dataset by giving the text as a separate modality for the shared task because the text linked with the meme acts as a context for the image. Both text and image-based classification approach, that is, multimodal classification, was expected from the participants.

Split	Troll	Not-troll	Total
Train	1,282	1,018	2,300
Test	395	272	667
Total	1,677	1,290	2,967

Table 3: Number of images in each class.

Team	Tamil	Rank
cuet_nlp_undergrad	0.24	1
baseline	0.20	2

Table 4: Macro-average F1-score of submitted tasks (for Tamil videos)

3.1 Dataset Description

The dataset consists of images and text associated with a meme in Tamil. The data is labelled into two classes, namely "troll" or "not-troll". The text in the meme is written in Tamil grammar with English lexicon or English grammar with Tamil lexicon. But for consistency the text was transcripted to Latin (Suryawanshi and Chakravarthi, 2021b), (Suryawanshi et al., 2020b).

- **Troll**: The meme has offensive text and non-offensive images, offensive images and non-offensive text, sarcastically offensive text with non-offensive images, or sarcastic images with offensive text to distract, distract, and digressive or off-topic content with the intent to demean or offend particular people, groups or race.

- **Not-Troll**: Images that do not have the above features are Non-Troll memes.

Table 3 gives the various classes and statistics of data points in each class and split up into train and test sets.

4 Methodology

4.1 Multimodal Sentiment Analysis

Multimodal Sentiment Analysis in Dravidian languages is organized for the first time. The task is challenging as the participants have to consider features extracted from three modalities to build a machine learning or deep learning model. Moreover, the size of the dataset in Malayalam and Tamil are 70 and 64, respectively, which is not sufficient to build a machine learning model. Hence, we received only one submission for this task. The rank list for this task is shown in Table 4.

Team	F1-score	Rank
BPHC	0.596	1
hate-alert	0.561	2
SSN_MLRG1	0.558	3
CUET89109115	0.529	4
DLRG_RR	0.519	5
TeamX	0.466	6

Table 5: The Weighted-average F1-score of teams submitted their predictions in Troll meme classification in Tamil

4.2 Troll meme Classification

A meme, also known as a troll, is an image with obscene or satirical text to degrade, provoke, or offend a person or a group. A troll meme can also be an image without any words. This section summarises the shared task for detecting troll memes using images and descriptions submitted by various participants. The main goal of this task is to classify trolling from multimodal memes. While reading the articles, we see that three of the five works submitted employ VGG16, a CNN-based design. Transformer-based models have also been proposed for extracting features from the text modal. One of the submitted tasks has combined the embeddings from text and image-based models. Below, we present a summary of each of the work submitted for this shared task, and the rank list of the task is shown in Table 5.

- TeamX (Rabindra Nath et al., 2022) used TF-IDF for extracting text features and SVM for training and classification of text data. In addition, multilingual BERT has been used for text classification. BERT model has also been trained to extract text from memes using Tesseract (Smith, 2007) and memes in hateful meme dataset (Kiela et al., 2021). Similarly, for image modality, pre-trained models such as EffNet, VGG16, and ResNet have been developed for detecting troll images. For text-only modality, mBERT with code mixed data has the highest accuracy among other models. EffNet has given better accuracy when compared to VGG16 and ResNet for image modality.

- A work by team hate-alert (Mithun et al., 2022) proposed two uni-modal models, one utilizing text features and the other using image-based features. All the texts associated with the meme are sent to a transformer model, MURIL, to get the dimensional feature vectors for each meme and then fed to an output node for the final prediction. For extracting features from images, VGG16 has been used. All the images have been passed to VGG16 and obtained feature vectors. Then, the embeddings from both MURIL and VGG16 models have been concatenated and sent to a classification node for the final prediction. The authors demonstrated that concatenated models give better accuracy than uni-modal models.

- Shruthi et al. (Shruthi et al., 2022) classified memes using three models: BERT, ALBERT, and XLNet. The first model is BERT, which has learned the context of a word based on all of its surroundings. Following this model, two more models, namely ALBERT (a Lite BERT) and XLNet, have also been utilized. The accuracy obtained by the XLNet model is 0.59, which is higher than BERT and ALBERT.

- An attempt by (Achyuta Krishna and Mithun Kumar, 2022) used two techniques to obtain embeddings from raw Tamil-English code-mixed text and another from translated and transliterated version of the dataset. The first approach used TF-IDF, LSTM and mBERT to retrieve embeddings. TF-IDF, IndicFT, MuRIL and mBERT have been used for getting embeddings in the second approach. Then, different machine learning-based classification algorithms, such as Naïve Bayes, Logistic Regression etc., have been used for detecting troll memes. Comparing the various pre-trained models, MuRIL performs best with an F1-score of 0.74 relative to others.

- Team CUET-NLP (Md Maruf et al., 2022) extracted the visual features of memes using CNN architectures such as VGG16, VGG19 and ResNet with transfer learning. Then, to extract the textual features, CNN and CNN with LSTM have been used due to their effectiveness in capturing the long-term dependencies from the long text. Subsequently, the output from the visual and textual models has been concatenated to form a single integrated model. The overall performance of the models is between 44 and 56% weighted F1-score.

From the articles submitted under this shared

task, we find that CNN based VGG models have been the choice of the authors for extracting the visual features from the troll memes and transformer-based models for extracting textual features.

5 Conclusion

We presented two multimodal shared tasks - Multimodal Sentiment Analysis and Troll meme classification in Dravidian languages. Multimodal Sentiment Analysis in Dravidian languages is the first shared task in this area. In addition to the multimodality, the code-mixing nature of the language posed challenges in these two tasks. We created datasets for both tasks to aid the research in the under-researched area. The machine models used by the participants revealed different ways of solving the problems mentioned above. We received one submission in the Multimodal Sentiment Analysis task and six submissions in Troll meme classification.

Acknowledgments

References

V Achyuta Krishna and S R Mithun Kumar. 2022. BPHC@DravidianLangTech-ACL2022: A Comparative Analysis of Classical and Pre-trained Models for Troll Meme Classification in Tamil Code Mixed Captions. In *Proceedings of the Second Workshop on Speech and Language Technologies for Dravidian Languages*. Association for Computational Linguistics.

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.

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.

Santiago Castro, Devamanyu Hazarika, Verónica Pérez-Rosas, Roger Zimmermann, Rada Mihalcea, and Soujanya Poria. 2019. Towards multimodal sarcasm detection (an _obviously_ perfect paper). *arXiv preprint arXiv:1906.01815*.

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, Barcelona, Spain (Online). Association for Computational Linguistics.

Bharathi Raja Chakravarthi, P K Jishnu Parameswaran, B Premjith, KP Soman, Rahul Ponnusamy, Prasanna Kumar Kumaresan, Kingston Pal Thamburaj, John P McCrae, et al. 2021a. Dravidian-multimodality: A dataset for multi-modal sentiment analysis in tamil and malayalam. *arXiv preprint arXiv:2106.04853*.

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, 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, Sajeetha Thavareesan, Dhivya Chinnappa, Durairaj Thenmozhi, Elizabeth Sherly, John P McCrae, Adeep Hande, Rahul Ponnusamy, Shubhanker Banerjee, et al. 2021b. Findings of the sentiment analysis of dravidian languages in code-mixed text. *arXiv preprint arXiv:2111.09811*.

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.

Douwe Kiela, Hamed Firooz, Aravind Mohan, Vedanuj Goswami, Amanpreet Singh, Casey A Fitzpatrick, Peter Bull, Greg Lipstein, Tony Nelli, Ron Zhu, et al. 2021. The hateful memes challenge: competition report. In *NeurIPS 2020 Competition and Demonstration Track*, pages 344–360. PMLR.

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.

Hasan Md Maruf, Jannat Nusratul, Hossain Eftekhar, Sharif Omar, and Hoque Mohammed Moshiul. 2022. CUET-NLP@DravidianLangTech-ACL2022: Investigating Deep Learning Techniques to Detect Multimodal Troll Memes. In *Proceedings of the Second Workshop on Speech and Language Technologies for Dravidian Languages*. Association for Computational Linguistics.

Das Mithun, Banerjee Somnath, and Mukherjee Animesh. 2022. hate-alert@DravidianLangTech-ACL2022: Ensembling Multi-Modalities for Tamil TrollMeme Classification. In *Proceedings of the Second Workshop on Speech and Language Technologies for Dravidian Languages*. Association for Computational Linguistics.

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.

B Premjith and KP Soman. 2021. Deep learning approach for the morphological synthesis in malayalam and tamil at the character level. *Transactions on Asian and Low-Resource Language Information Processing*, 20(6):1–17.

B Premjith, KP Soman, M Anand Kumar, and D Jyothi Ratnam. 2019. Embedding linguistic features in word embedding for preposition sense disambiguation in english—malayalam machine translation context. In *Recent Advances in Computational Intelligence*, pages 341–370. Springer.

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.

Zea Qiyang and Heekyoung Jung. 2019. Learning and sharing creative skills with short videos: A case study of user behavior in tiktok and bilibili. *International association of societies of design research (IASDR), design revolution*.

Nandi Rabindra Nath, Alam Firoj, and Nakov Preslav. 2022. TeamX@DravidianLangTech-ACL2022: A Comparative Analysis for Troll-Based Meme Classification. In *Proceedings of the Second Workshop on Speech and Language Technologies for Dravidian Languages*. Association for Computational Linguistics.

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.

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.

Tobias Schreck and Daniel Keim. 2012. Visual analysis of social media data. *Computer*, 46(5):68–75.

H Shruthi, Esackimuthu Sarika, M Saritha, S Rajalakshmi, and S Angel Deborah. 2022. SSN_MLRG1@DravidianLangTech-ACL2022: Troll Meme Classification in Tamil using Transformer Models. In *Proceedings of the Second*

Workshop on Speech and Language Technologies for Dravidian Languages. Association for Computational Linguistics.

Ray Smith. 2007. An overview of the tesseract ocr engine. In *Ninth international conference on document analysis and recognition (ICDAR 2007)*, volume 2, pages 629–633. IEEE.

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.

Shardul Suryawanshi and Bharathi Raja Chakravarthi. 2021a. Findings of the shared task on troll meme classification in tamil. In *Proceedings of the First Workshop on Speech and Language Technologies for Dravidian Languages*, pages 126–132.

Shardul Suryawanshi and Bharathi Raja Chakravarthi. 2021b. Findings of the shared task on troll meme classification in tamil. In *Proceedings of the First Workshop on Speech and Language Technologies for Dravidian Languages*, pages 126–132.

Shardul Suryawanshi, Bharathi Raja Chakravarthi, Pranav Verma, Mihael Arcan, John Philip McCrae, and Paul Buitelaar. 2020a. A dataset for troll classification of tamilmemes. In *Proceedings of the WILDRE5–5th workshop on indian language data: resources and evaluation*, pages 7–13.

Shardul Suryawanshi, Bharathi Raja Chakravarthi, Pranav Verma, Mihael Arcan, John Philip McCrae, and Paul Buitelaar. 2020b. A dataset for troll classification of tamilmemes. In *Proceedings of the WILDRE5–5th workshop on indian language data: resources and evaluation*, pages 7–13.

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.

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.