Aanisha@TamilNLP-ACL2022:Abusive Detection in Tamil

Aanisha Bhattacharyya
Institute of Engineering and Management
aanishabhattacharyya@gmail.com

Abstract

In social media, there are instances where people present their opinions in strong language, resorting to abusive/toxic comments.There are instances of communal hatred, hate-speech, toxicity and bullying. And, in this age of social media, it's very important to find means to keep check on these toxic comments, as to preserve the mental peace of people in social media.While there are tools, models to detect and potentially filter these kind of content, developing these kinds of models for the low resource language space is an issue of research.

In this paper, the task of abusive comment identification in Tamil language, is seen upon as a multiclass classification problem.There are different pre-processing as well as modelling approaches discussed in this paper.The different approaches are compared on the basis of weighted average accuracy.

[1]

1 Introduction

With social media being accessible and popular across masses in India, there has been a surge in content in regional languages. People often create content, comment or exchange messages in monolingual or code mixed language (Priyadharshini et al., 2020, 2021; Kumaresan et al., 2021). However,even if there is an abundance of content in Indian language across social media, there is a lack of Indian language datasets (Chakravarthi, 2020; Chakravarthi and Muralidaran, 2021). Hence Indian languages are deemed as low resource language space, due to lack of available datasets, making working in these languages spaces, a challenging research problem (Chakravarthi et al., 2019b, 2018).

Among the messages and comments exchanged on social media there are instances of monolingual comments in regional language as well as transliterated comments. Monolingual comments in transliterated means to write or print (a letter or word) using the closest corresponding letters of a different alphabet or script. Code-Mixing is mixing of two or more language in the same utterance (Sampath et al., 2022; Ravikiran et al., 2022; Chakravarthi et al., 2022a; Bharathi et al., 2022; Priyadharshini et al., 2022).

In this paper, the task is identifying abusive comments in Tamil language. Tamil is a member of the southern branch of the Dravidian languages, a group of about 26 languages indigenous to the Indian subcontinent (Anita and Subalalitha, 2019b,a; Subalalitha and Poovammal, 2018). It is also classed as a member of the Tamil language family, which contains the languages of around 35 ethno-linguistic groups, including the Irula and Yerukula languages (Subalalitha, 2019; Srinivasan and Subalalitha, 2019; Narasimhan et al., 2018). The earliest Old Tamil documents are small inscriptions in Adichanallur dating from 905 BC to 696 BC. This is a multiclass classification problem, with 6 different categories of abusive comments are present. In a multi class classification problem, an instance can belong only to one class. However present Machine Learning or Deep Learning based models cannot be directly applied to Tamil language. Thus several pre-processing techniques have been proposed for Tamil language and models have been fine tuned to suit the task (Sakuntharaj and Mahesan, 2021, 2017, 2016; Thavareesan and Mahesan, 2019, 2020a,b, 2021).

2 Related work

There has been different works done on identifying abusive comments on different languages.

In (Zhao et al., 2021) Performed binary and multiclass classification using a Twitter corpus and studied two approaches: (a)a method which consists in extracting of word embeddings and then using a DNN classifier; (b) fine-tuning the pre-trained

[1]https://github.com/Aanisha/Tamil_Comment_Classification

Proceedings of the Second Workshop on Speech and Language Technologies for Dravidian Languages, pages 214 - 220

BERT model.However it was only on English language embeddings.

In (Farooqi et al., 2021) Detected hate speech from Hindi-English code mixed conversations on Twitter. The proposed architecture used neural networks, leveraging the transformer's cross-lingual embeddings and further fine tuning them for low-resource hate-speech classification in transliterated Hindi text.

In (Andrew, 2021) as a part of shared task in ACL Dravidian Lang Tech 2021, several Machine learning algorithms were compared and experimented for identifying abusive comment in various Dravidian languages.

3 Dataset

The dataset is provided by (Priyadharshini et al., 2022) as a part of the shared task Abusive comment detection in Tamil.

The dataset has a collection of comments in Tamil language.There are 2240 native Tamil script comments and 5943 transliterated Tamil-English comments in the train data,classified across 7 different categories : 'Hope-Speech', 'Homophobia', 'Misandry', 'Counter-speech', 'Misogyny', 'Xenophobia', Trans-phobic' and 'None-of-the-above'.

The validation data has 560 native Tamil script comments and 1486 transliterated Tamil-English comments.The test data has 699 native Tamil language comments and 1857 transliterated Tamil-English comments.

The most dominant category present across all the datasets is : 'None-of-the-above' and the categories with less no of comments are 'Homophobia' with 207 and 'Trans-phobic' with 163 total comments.

4 Approaches

4.1 Pre-processing

The dataset has a very imbalanced distribution of the categories of comments.

So, for the experiments two separate datasets are generated.

Table 1 shows the distribution of first dataset, is combining both native Tamil script and transliterated Tamil-English comments.

Table 2, shows the distribution of second dataset,which creates a more balanced distribution by a mixed approach of oversampling and undersampling.

Command	**Output**
`None-of-the-above`	5011
`Misandry`	1276
`Counter-speech`	497
`Xenophobia`	392
`Misogyny`	336
`Hope-Speech`	299
`Homophobia`	207
`Trans-phobic`	163

Table 1: Distribution of comments in the different categories

Command	**Output**
`None-of-the-above`	3007
`Misandry`	1276
`Counter-speech`	497
`Xenophobia`	392
`Misogyny`	586
`Hope-Speech`	549
`Homophobia`	457
`Trans-phobic`	413

Table 2: Distribution of comments in the different categories in pre-processed dataset.

The 'None-of-the-above' class comments are downsampled by a percentage of 0.4 in the train data .The lower represented classes 'Misogyny','Hope-speech','Homophobia' and 'Trans-phobic' data samples are over-sampled.

The values are decided on experimental basis.

4.2 Tokenization and feature vectors

For tokenization of the dataset, two different tokenizers have been used.

The MuRil tokenizer is used.MuRIL is a multilingual LMBert specifically built for IN languages.MuRIL is trained on significantly large amounts of IN text corpora only.Can generate embeddings for low resource native script and transliterated Indic languages.(Khanuja et al., 2021).

Another tokenizer used is the the IndicNLP tokenizer. A trivial tokenizer which just tokenizes on the punctuation boundaries. This also includes punctuations for the Indian language scripts (the purna virama and the deergha virama). It returns a list of tokens.(Arora, 2020).

Two kinds of feature vectors are used for the various modelling approaches.MuRil embeddings are generated from pre-trained MuRil model, and used as feature vectors for solving the multiclass classification problem.

Another feature vector used is normalized Tf-Idf vectors from the tokenized text,where

tf(t) = (No. of times term 't' occurs in a document) / (Frequency of most common term in a document)

and,

idf(t) = log e [(1+Total number of documents available) / (1 + Number of documents in which the term t appears)] + 1)

$$tf - idf(t) = tf(t) * idf(t)$$

These feature vectors are generated from the tokenized texts of MuRil and IndicNLP tokenizer respectively.

4.3 Modelling approaches

4.3.1 Logistic Regression

The multiclass logistic regression model is implemented (LR, 2017). The model of logistic regression for a multiclass classification problem forces the output layer to have discrete probability distributions over the possible k classes. This is accomplished by using the softmax function. Given the input vector(z), the softmax function works as follows:

$\sigma(z_i) = \frac{e^{z_i}}{\sum_{j=1}^{K} e^{z_j}} \quad for\ i = 1, 2, \ldots, K$

There are n output classes and thus there is a necessity to impose weights connecting each input to each output.

4.3.2 Linear Support Vector Machines

SVMs are very good classification algorithm. The idea is to identify hyper-planes that will separate the various features. The classification decision is thus performed as follows:

$f(x) = sign(W.x + b)$

where x represents the input feature, W represents the model weight and b represents the bias. For the multiclass classification problem, a one-vs-rest (also known as one-vs-all) approach is used.

4.3.3 Gradient Boosting Classifier

Gradient boosting classifiers are a group of machine learning algorithms that combine many weak learning models together to create a strong predictive model. Decision trees are usually used when doing gradient boosting.

Here, this algorithm is used for a multiclass classification.

4.3.4 Transformers

Google introduced the transformer architecture in the paper "Attention is All you need". Transformer uses a self-attention mechanism, which is suitable for language understanding.The transformer has an encoder-decoder architecture. They are composed of modules that contain feed-forward and attention layers.

They have led to advancements in the field of NLP to perform tasks as text classification,machine translation etc.

5 Results

6 Implementation

6.0.1 Logistic Regression

The original training data contains 10227 comments and the test data contains 2555 comments.

The data is first tokenized using the IndicNLP tokenizer and feature vectors are generated by using Tf-Idf with unigrams and bigrams being extracted.

The feature vector are fed to the logistic regression model with a newton-cg solver, to accomodate multiclass classification.

There are two experiments that are run for this model.The model is trained on original dataset and the model is trained on sampled dataset.

6.0.2 Support Vector Machine

The original training data contains 10227 comments and the test data contains 2555 comments.

The data is first tokenized using the IndicNLP tokenizer and feature vectors are generated by using Tf-Idf with unigrams and bigrams being extracted.

The feature vector are fed to the support vector machine with degree=8, to accomodate multiclass classification.The penalty is squared l2.

There are two experiments that are run for this model.The model is trained on original dataset and the model is trained on sampled dataset.

6.0.3 Gradient Boosting Classifier

The original training data contains 10227 comments and the test data contains 2555 comments.

The data is first tokenized using the IndicNLP tokenizer and feature vectors are generated by using Tf-Idf with unigrams and bigrams being extracted.

The feature vector is input to the Gradient Boosting Classifier model, which uses deviance loss function for optimization and a learning rate of 0.1.

Model	Dataset	Acc	Precision	Recall	F1-score
`Logistic Regression`	`Original`	0.66	0.62	0.66	0.57
`Logistic Regression`	`Sampled`	0.65	0.62	0.65	0.59
`Linear SVM`	`Original`	0.59	0.54	0.59	0.47
`Linear SVM`	`Sampled`	0.56	0.50	0.56	0.48
`Gradient Boost Classifier`	`Original`	0.68	0.67	0.68	0.63
`Gradient Boost Classifier`	`Sampled`	0.70	0.67	0.70	0.66
`Finetuned MuRIL`	`Original`	0.68	0.60	0.68	0.62
`Finetuned MuRIL`	`Sampled`	0.64	0.67	0.64	0.65
`Finetuned MuRIL(weighted loss)`	`Sampled`	0.51	0.67	0.51	0.56

Table 3: The results of the experiments conducted.

There are two experiments that are run for this model.The model is trained on original dataset and the model is trained on sampled dataset.

6.0.4 Transformers

The train data contains 8183 comments and the validation data contains 2046 comments.Also the sampled train dataset(details in dataset) is tested on this system.Validation data is same in both the experiments.

The data is tokenised using the Muril tokeniser which has a a vocabulary of 197,285.

The tokenised output from the MuRil tokenizer has 3 elements `Input Id`,`Attention Mask` and `Token Id`.These 3 vectors are fed to the pre-trained MuRil model to generate embeddings.

The model embeddings are input to a 1D convolutional layer which changes the dimension of the embedding from (x,64,768) to (x,64,1).Then it's flattened to have a vector of dimension (x,64).Lastly, there is a fully connected layer with softmax activation to have the output of dim. (x,8).The model output is the probabilities for the sentence to belong to each of the categories.

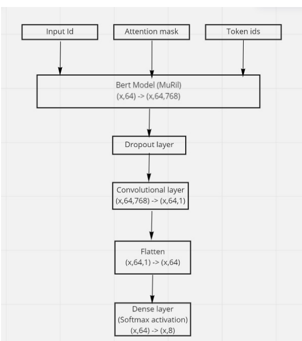

Figure 1: The finetuned MuRil model

For training, the MuRil layers are frozen and pre-trained weights are used.Only trainable layers are the CNN and Dense layers.There is a dropout of 0.2 used.

There are three experiments that are run for this model.The model is trained on original dataset,the model is trained on sampled dataset, and the model is trained on sampled dataset with weighted loss being applied.

The models in each case are trained for 25 epochs.All the transformers are trained on a single GPU and takes around 25-30 mins for one training session.

7 Results

Table 3 contains the results from the different experiments. The best performing model is the Gradient Boost Classifier trained on the sampled dataset. Within the results, the category "None-of-the-above" is more easily detected correctly by most of the models, while the classes "Misogyny" and "counter-speech" are not detected easily. The transformer finetuned on original dataset has the highest accuracy among all the transformer experiments. However it's not able to identify the categories with lower number of datapoints. The transformers trained on sampled dataset is able to perform better in the categories with lower number of datapoints.

8 Future Work

The future work will be primarily to find more efficient sampling techniques for the text data, and compare the performances with further ML models. Also, evaluate performances with other existing transformer models,to check how different suitable models can be fine-tuned to solve this particular task.

9 References

References

BERT and fastText Embeddings for Automatic Detection of Toxic Speech.

Judith Jeyafreeda Andrew. 2021. JudithJeyafreedaAndrew@DravidianLangTech-EACL2021:offensive language detection for Dravidian code-mixed YouTube comments. In *Proceedings of the First Workshop on Speech and Language Technologies for Dravidian Languages*, pages 169–174, Kyiv. 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.

Gaurav Arora. 2020. *iNLTK: Natural Language Toolkit for Indic Languages*.

Bharathi B and Agnusimmaculate Silvia A. 2021a. SSNCSE_NLP@DravidianLangTech-EACL2021: Meme classification for Tamil using machine learning approach. In *Proceedings of the First Workshop on Speech and Language Technologies for Dravidian Languages*, pages 336–339, Kyiv. Association for Computational Linguistics.

Bharathi B and Agnusimmaculate Silvia A. 2021b. SSNCSE_NLP@DravidianLangTech-EACL2021: Offensive language identification on multilingual code mixing text. In *Proceedings of the First Workshop on Speech and Language Technologies for Dravidian Languages*, pages 313–318, Kyiv. Association for Computational Linguistics.

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

Bharathi Raja Chakravarthi, Mihael Arcan, and John P. McCrae. 2018. Improving wordnets for under-resourced languages using machine translation. In *Proceedings of the 9th Global Wordnet Conference*, pages 77–86, Nanyang Technological University (NTU), Singapore. Global Wordnet Association.

Bharathi Raja Chakravarthi, Mihael Arcan, and John P McCrae. 2019a. Comparison of different orthographies for machine translation of under-resourced Dravidian languages. In *2nd Conference on Language, Data and Knowledge (LDK 2019)*. Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik.

Bharathi Raja Chakravarthi, Mihael Arcan, and John P. McCrae. 2019b. WordNet gloss translation for under-resourced languages using multilingual neural machine translation. In *Proceedings of the Second Workshop on Multilingualism at the Intersection of Knowledge Bases and Machine Translation*, pages 1–7, Dublin, Ireland. European Association for Machine Translation.

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. 2022a. 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, Navya Jose, Shardul Suryawanshi, Elizabeth Sherly, and John P. McCrae. 2022b. DravidianCodeMix: sentiment analysis and offensive language identification dataset for Dravidian languages in code-mixed text. *Language Resources and Evaluation*.

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. 2021a. Dataset for identification of homophobia and transophobia in multilingual YouTube comments. *arXiv preprint arXiv:2109.00227*.

Bharathi Raja Chakravarthi, Ruba Priyadharshini, Bernardo Stearns, Arun Jayapal, Sridevy S, Mihael Arcan, Manel Zarrouk, and John P McCrae. 2019c. Multilingual multimodal machine translation for Dravidian languages utilizing phonetic transcription. In *Proceedings of the 2nd Workshop on Technologies for MT of Low Resource Languages*, pages 56–63, Dublin, Ireland. European Association for Machine Translation.

Bharathi Raja Chakravarthi, Priya Rani, Mihael Arcan, and John P McCrae. 2021b. A survey of orthographic information in machine translation. *SN Computer Science*, 2(4):1–19.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. *BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding*.

Zaki Mustafa Farooqi, Sreyan Ghosh, and Rajiv Ratn Shah. 2021. *Leveraging Transformers for Hate Speech Detection in Conversational Code-Mixed Tweets*.

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.

Priyanka Gupta, Shriya Gandhi, and Bharathi Raja Chakravarthi. 2021. Leveraging transfer learning techniques-BERT, RoBERTa, ALBERT and DistilBERT for fake review detection. In *Forum for Information Retrieval Evaluation*, pages 75–82.

Simran Khanuja, Diksha Bansal, Sarvesh Mehtani, Savya Khosla, Atreyee Dey, and Pooja Aggarwal Rajiv Teja Nagipogu Shachi Dave Shruti Gupta Subhash Chandra Bose Gali Vish Subramanian Partha Talukdar Balaji Gopalan, Dilip Kumar Margam. 2021. *MuRIL:Multilingual Representations for Indian Languages*.

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.

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.

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.

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.

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.

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.

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.

Zhixue Zhao, Ziqi Zhang, and Frank Hopfgartner. 2021. *A Comparative Study of Using Pre-trained Language Models for Toxic Comment Classification*.