IIITSurat@LT-EDI-ACL2022: Hope Speech Detection using Machine Learning

Pradeep Kumar Roy[1], **Snehaan Bhawal**[2], **Abhinav Kumar**[3], and **Bharathi Raja Chakravarthi**[4]
[1]Indian Institute of Information Technology Surat, Gujarat, India
[2]Kalinga Institute of Industrial Technology, Bhubaneswar, Odisha, India
[3]Siksha 'O' Anusandhan, Deemed to be University, Bhubanewar, Odisha, India
[4]Insight SFI Research Centre for Data Analytics, National University of Ireland Galway, Ireland
(pkroynitp,mailtosnehaan,abhinavanand05)@gmail.com,
bharathi.raja@insight-centre.org

Abstract

This paper addresses the issue of Hope Speech detection using machine learning techniques. Designing a robust model that helps in predicting the target class with higher accuracy is a challenging task in machine learning, especially when the distribution of the class labels is highly imbalanced. This study uses and compares the experimental outcomes of the different oversampling techniques. Many models are implemented to classify the comments into Hope and Non-Hope speech, and it found that machine learning algorithms perform better than deep learning models. The English language dataset used in this research was developed by collecting YouTube comments and is part of the task "ACL-2022:Hope Speech Detection for Equality, Diversity, and Inclusion". The proposed model achieved a weighted F1-score of 0.55 on the test dataset and secured the first rank among the participated teams.

1 Introduction

Social networking platforms such as Instagram, Facebook, LinkedIn, and YouTube have become the default place for worldwide users to spend time (Chakravarthi et al., 2021, 2020; Priyadharshini et al., 2020). These social platforms are not only used to share success but also used to ask for help during emergency (Roy et al., 2021). As per the report[1], on average, six hours in a week, every Indian uses the social networking platform. Among them, teenagers and some professionals are more active to share their life events.

People have two images: one for the real world where they live and another for the virtual world, like the images on social platforms where people are connected to their close friends and communicate with strangers in the virtual environment (Saumya and Mishra, 2021). Language is a primary requirement for communication. Languages like Hindi, English, Japanese, Gujarati, Marathi, Tamil, and others are used to express success, life events like job promotion, being selected as the best team member, etc (Sampath et al., 2022; Ravikiran et al., 2022; Chakravarthi et al., 2022b; Bharathi et al., 2022; Priyadharshini et al., 2022). Tamil is one of the world's longest-surviving classical languages. Tamil is a member of the southern branch of the Dravidian languages, a group of about 26 languages indigenous to the Indian subcontinent. 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 (Sakuntharaj and Mahesan, 2021, 2017, 2016; Thavareesan and Mahesan, 2019, 2020a,b, 2021). The earliest Old Tamil documents are small inscriptions in Adichanallur dating from 905 BC to 696 BC. Tamil has the oldest ancient non-Sanskritic Indian literature of any Indian language (Anita and Subalalitha, 2019b,a; Subalalitha and Poovammal, 2018; Subalalitha, 2019; Srinivasan and Subalalitha, 2019; Narasimhan et al., 2018).

Everyone needs feelings like happiness, sadness, anger, and the motivation for failure in their hard time (Ghanghor et al., 2021b; Yasaswini et al., 2021). Among all, the comments having the context of "well-being" are termed as "hope speech". More specifically, Hope speech reflects the belief that one can discover and become motivated to use pathways to achieve one's desired goals (Chang, 1998; Youssef and Luthans, 2007; Cover, 2013; Snyder et al., 1991). The other category of comments can be abuse, demotivate, neutral, race, or sexually oriented and similar ones which are termed as "Non-Hope speech". Such comments do not live long in the physical world where people speak something today that might not be remembered after a few days or months, even the reachable to the limited region. However, if the same is communicated via a social platform, it will re-

[1]https://www.statista.com/statistics/1241323/

Proceedings of the Second Workshop on Language Technology for Equality, Diversity and Inclusion, pages 120 - 126
May 27, 2022 ©2022 Association for Computational Linguistics

main active and affect the victim for a long-time (Saumya and Mishra, 2021).

The social platform is polluted with hateful content (Roy et al., 2020) and is a challenging task to filter. Moreover, finding the hopeful message becomes another challenging task because of their low appearance. People who are in trouble are expecting a solution for their issues. For example, if a person becomes a victim of cybercrime like borrowing money from a bank account. Then they will reach out to the concerned authority hoping that their money will be rolled back into the account. If people face issues with the company rules and regulations, they ask for opinions via social posts hoping that someone will suggest the right solution to get rid of it.

These social platforms receive huge content from worldwide users from different genres like entertainment, promotion, publicity, achievement, political news, etc. Every genre has both positive and negative comments. All of the mentioned scenarios are common in human life, where directly or indirectly, people always expect some positive news with hope (Chakravarthi, 2020). Finding hope speech content from social platforms manually is challenging and not a feasible option. Hence there is a need of automated tools which can be helpful for hope-oriented comment detection (Chakravarthi and Muralidaran, 2021; Chakravarthi et al., 2022a). To address the said problem, this research uses both traditional machine learning (ML) models and deep learning (DL) based models to find the best-suited technique to detect such hope speech. The dataset used in this research was taken from LT-EDI-ACL2022 workshop. The major contributions are as follows:

- We proposed an automated machine learning-based model to predict hope speech.
- Performed data balancing techniques to balance the samples in each category.
- The machine learning model outperformed deep learning models on a balanced dataset.

The rest of the paper is organized as follows: Section 2 discusses the relevant research works. Section 3 describes the overview of the task in detail. Section 4 explains the data preparation for the experiment followed by experimental setup in Section 5. Section 6 discusses the experimental outcomes of different models. Finally, the work is concluded in Section 7 with limitation and future scope.

2 Related works

Even though the Hope speech is termed as positive vibes, very less attention is received from the research community to address it. The reason behind less research in the domain may include the unavailability of the labeled dataset. In the last few years, this problem has received some fruitful attention while the organizer of the LT-EDI-EACL2021 shared a labeled dataset. Some of the submitted frameworks in the LT-EDI-EACL2021 workshop is to address this Hope Speech detection issue. Many research works have reported to filter the Hateful, and Offensive comments from the social post in recent years (Roy et al., 2022; Ghanghor et al., 2021a). However, identifying the Hopeful comments received less attention (Chakravarthi, 2020; Hande et al., 2021; Saumya and Mishra, 2021).

(Puranik et al., 2021) used transformer-based models like BERT, ALBERT, DistilBERT, and similar ones to classify the comments into three categories: hope, non-hope, and other categories. Dataset of three languages were used in their research, English, Malayalam, and Tamil. For the English language, the ULMFit model achieved the best weighted F1-score value of 0.9356. (Upadhyay et al., 2021) also used the transformer-based model to classify the comments into hope, non-hope, and other categories. Deep learning models - Convolutional Neural Network (CNN), Long Short Term Memory (LSMT), and Bidirectional LSTM approaches were used by (Saumya and Mishra, 2021) on all three datasets. Their best-performing CNN-LSTM model achieved an F1-score of 0.91 on English.

3 Task and Dataset Overview

In LT-EDI-ACL2022, Task 1 was Hope Speech Detection for Equality, Diversity, and Inclusion, where the event organizer provided an annotated dataset for three languages Tamil, Malayalam, and English. The dataset was labeled into two categories: 'Hope Speech and Non-Hope Speech'. The shared task's objective was to build an automated model that predicts the comments are Hope Speech or Non-Hope Speech. Initially, the training dataset was released. Later, the validation and test dataset was released by the organizer. This research uses only English comments for the experiment. The training

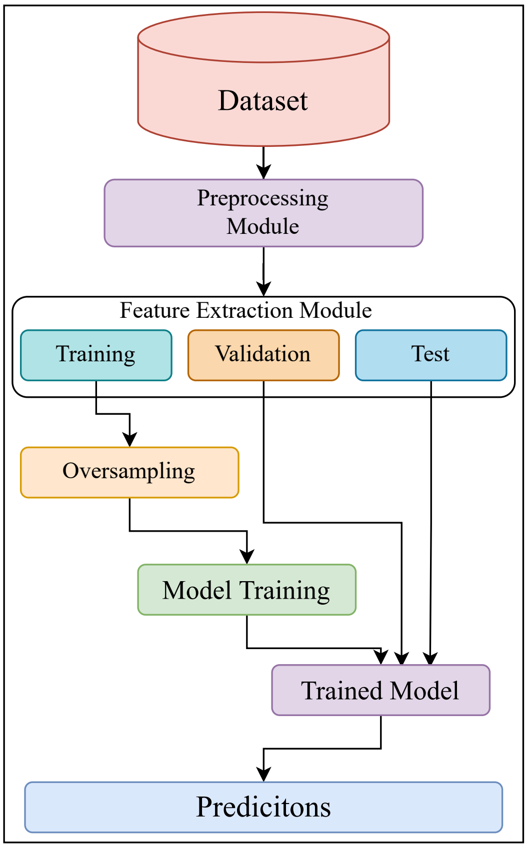

Figure 1: Working steps of the proposed model

dataset had a total of 20778 numbers of Non-Hope Speech sample whereas in Hope speech 1962 sample. 2569 Non-Hope Speech and 272 Hope Speech samples were present in the validation dataset. Finally, the test dataset was released without any label on which the final rank of the participated teams was decided (Chakravarthi and Muralidaran, 2021; Chakravarthi, 2020; Hande et al., 2021).

4 Data Preprocessing

As the dataset was compiled with comments collected from YouTube, it consisted of many irregularities like the use of emoticons/emojis, short text, customized fonts, and tagged users. All these need to be cleaned for the data to be passed onto the model for training. During the preprocessing of the data, the emojis were replaced with their mapped meaning by using Demoji library[2]. Tagged users and punctuation were removed and also removed all custom fonts and numerals, single-character words, and multiple spaces that were introduced by the previous steps.

[2]https://pypi.org/project/demoji/

Table 1: Label Distribution of the dataset

Data Set	Hope	Non-Hope	Total
Train	1,962	20,778	22,740
Validation	272	2,569	2,841

Table 2: Average accuracy obtained using ML classifiers on different data balancing approaches (No oversampling (NO), Random Oversampling (ROS), SMOTE and ADASYN)

Model	NO	ROS	SMOTE	ADASYN
LR	0.926	0.920	0.921	0.893
RF	0.925	0.992	0.971	0.962
NB	0.915	0.848	0.866	0.836
XGB	0.924	0.910	0.939	0.928

4.1 Oversampling

The dataset used for this research is highly imbalanced. The class-wise distribution of the dataset is shown in Table 1. The imbalanced dataset could lead to a biased model, and thus it is needed to balance the distribution of the class labels by oversampling the minority class. To make the dataset of both the classes comparable in the training sample, three oversampling techniques are used; namely, Random Oversampling (ROS) (Menardi and Torelli, 2014), Synthetic Minority Over-sampling Technique (SMOTE) (Chawla et al., 2002) and Adaptive Synthetic (ADASYN) (He et al., 2008). After oversampling, in both classes, the number of samples is 20778. Overall working steps of the proposed framework are shown in Figure 1.

5 Experimental setup

This section discusses a detailed experimental procedure used for the model development. The traditional ML techniques, namely, Logistic Regression (LR), Random Forest (RF), Naive Bayes (NB), and Extreme Gradient Boosting (XGB), are selected for the experiment. The performance of these models is evaluated with Precision, Recall, and F1 score (Roy et al., 2022). Firstly, a total of 5000 features were extracted from the processed data using TF-IDF vectorization with 1-5 n-grams, which was further scaled using the MIN-MAX scalar. The oversampling techniques mentioned above were used to balance the dataset before passing it to the model. Before oversampling, the total train data size was 22,740. After oversampling, the total number of samples increased to 41,556, with both the classes divided equally.

The balanced dataset was then passed to the ML classifiers with the help of 10-fold cross-validation over the training dataset. We implemented all the combinations of the selected classifiers and oversampling techniques. The average accuracy obtained using a 10-fold cross-validated approach is shown in Table 2. Based on these values, the SMOTE oversampling approach was selected for further experiments. The comparative outcomes of the ML classifiers on the imbalanced and balanced dataset are discussed in section 6.

Further, deep learning techniques like DNN, DNN with embeddings (DNN+Emb), CNN, LSTM, and BiLSTM are implemented to address this issue. The DNN model is comprised of a simple four-layer neural network with 256, 128, and 64 neurons at the hidden layer with a single output neuron. In DNN + Emb, we have implemented an additional embedding layer of 120 dimensions. A single convolution layer is used in CNN, followed by a MaxPooling layer and hidden layers of 128 and 64 neurons. Similarly, the LSTM and BiLSTM networks are implemented with 256 memory units with the same amounts of hidden layers. The output layer consisted of a single neuron with sigmoid activation for each model. After further hyperparameter tuning, we concluded by using the Adam optimizer with a learning rate of 0.0001 and binary cross-entropy as the optimization function. The model was trained with the SMOTE oversampled train data and was validated with the provided validation data set, the results of which are provided in Table 4.

6 Results

In this section, the experimental outcomes of the different models will be discussed. We are comparing the performances of the ML models based on the 10-fold cross-validated outcomes reported in Table 2. The table shows the average accuracy achieved by the individual models with the respective oversampling techniques used on the train data. The experimental outcomes when no oversampling ('NO'), i.e. the initial imbalanced dataset, was implemented in shown in Table 2. We can see that oversampling is not helpful for the NB and LR, where the performances are degraded in some cases. On the other hand, the RF model achieved the best performance with oversampled data. The performance of the XGB model remained consistent in all cases. This shows that model selection can vary with the oversampling technique chosen. The SMOTE technique is always outperforming the ADASYN while RF with ROS achieves 99% accuracy, which is probably due to the overfitting of the training samples. The validation data is used to validate the findings. Table 3 shows the results of the RF model with the different oversampling techniques. We can see that using Random Oversampling results in over-fitting. Thus, we chose SMOTE and Random Forest as the final model. The weighted F1-score of the RF model with a balanced dataset was compared with the deep learning techniques. The comparative outcomes are shown in Table 4. The RF model is performing better than the deep learning models.

Table 3: Detailed report of RF with different Oversampling techniques on validation data.

Model	**Class**	Precision	Recall	F1-score
	Hope	0.83	0.19	0.32
NO + RF	Non-Hope	0.92	1.00	0.96
	Weighted Avg	0.91	0.92	0.90
	Hope	0.78	0.26	0.39
ROS + RF	Non-Hope	0.93	0.99	0.96
	Weighted Avg	0.91	0.92	0.90
	Hope	0.64	0.41	0.50
SMOTE + RF	Non-Hope	0.94	0.98	0.96
	Weighted Avg	0.91	0.92	0.92

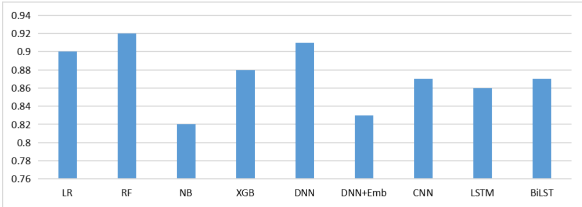

Figure 2: Comparison of F1-scores of experimented models on balanced data

Table 4: Comparison with the deep learning models

Model	F1-score
RF	**0.92**
DNN	0.91
DNN + Emb	0.83
CNN	0.87
LSTM	0.86
BiLSTM	0.87

7 Conclusion

Social platforms have become a medium to share opinions, achievements, successes, and failures.

Social networking users comment on all categories of posts. The comments having positive vibes is really help in boosting confidence and sometimes motivate to be strong in the odd situation. This paper suggested an ML model to predict the Hope Speech comments on the social platform. The samples available for training were highly imbalanced; hence, the SMOTE oversampling technique was used to balance the dataset. Many models have experimented on both imbalanced and balanced datasets, and it was found that the Random Forest classifier performed best when the training sample was balanced. The proposed balanced model secured top rank among the participated teams for the English language with a weighted F1-score of 0.550 on the test dataset. The model can be further tuned with preprocessing steps as well as by increasing the size of the feature set to achieve better performance. In the future, the transformer based model can be implemented, also ensemble models can be explored for the same in the future.

References

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.

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.

Bharathi Raja Chakravarthi, Vigneshwaran Muralidaran, Ruba Priyadharshini, Subalalitha Chinnaudayar Navaneethakrishnan, John Phillip McCrae, Miguel Ángel García-Cumbreras, Salud María Jiménez-Zafra, Rafael Valencia-García, Prasanna Kumar Kumaresan, Rahul Ponnusamy, Daniel García-Baena, and José Antonio García-Díaz. 2022a. Findings of the shared task on Hope Speech Detection for Equality, Diversity, and Inclusion. In *Proceedings of the Second Workshop on Language Technology for Equality, Diversity and Inclusion*. Association for Computational Linguistics.

Bharathi Raja Chakravarthi, Vigneshwaran Muralidaran, Ruba Priyadharshini, and John Philip McCrae. 2020. 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. 2022b. 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, 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*.

Edward C Chang. 1998. Hope, problem-solving ability, and coping in a college student population: Some implications for theory and practice. *Journal of Clinical Psychology*, 54(7):953–962.

Nitesh V Chawla, Kevin W Bowyer, Lawrence O Hall, and W Philip Kegelmeyer. 2002. Smote: synthetic minority over-sampling technique. *Journal of Artificial Intelligence Research*, 16:321–357.

Rob Cover. 2013. Queer youth resilience: Critiquing the discourse of hope and hopelessness in lgbt suicide representation. *M/C Journal*, 16(5).

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.

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. *arXiv preprint arXiv:2108.04616*.

Haibo He, Yang Bai, Edwardo A Garcia, and Shutao Li. 2008. Adasyn: Adaptive synthetic sampling approach for imbalanced learning. In *2008 IEEE international joint conference on neural networks (IEEE world congress on computational intelligence)*, pages 1322–1328. IEEE.

Giovanna Menardi and Nicola Torelli. 2014. Training and assessing classification rules with imbalanced data. *Data mining and knowledge discovery*, 28(1):92–122.

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, 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.

Karthik Puranik, Adeep Hande, Ruba Priyadharshini, Sajeetha Thavareesan, and Bharathi Raja Chakravarthi. 2021. IIITT@ LT-EDI-EACL2021-hope speech detection: there is always hope in transformers. *arXiv preprint arXiv:2104.09066*.

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.

Pradeep Kumar Roy, Snehaan Bhawal, and C.N. Subalalitha. 2022. Hate speech and offensive language detection in Dravidian languages using deep ensemble framework. *Computer Speech Language*, page 101386.

Pradeep Kumar Roy, Abhinav Kumar, Jyoti Prakash Singh, Yogesh Kumar Dwivedi, Nripendra Pratap Rana, and Ramakrishnan Raman. 2021. Disaster related social media content processing for sustainable cities. *Sustainable Cities and Society*, 75:103363.

Pradeep Kumar Roy, Asis Kumar Tripathy, Tapan Kumar Das, and Xiao-Zhi Gao. 2020. A framework for hate speech detection using deep convolutional neural network. *IEEE Access*, 8:204951–204962.

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.

Sunil Saumya and Ankit Kumar Mishra. 2021. IIIT_DWD@ LT-EDI-EACL2021: hope speech detection in YouTube multilingual comments. In *Proceedings of the First Workshop on Language Technology for Equality, Diversity and Inclusion*, pages 107–113.

Charles R Snyder, Cheri Harris, John R Anderson, Sharon A Holleran, Lori M Irving, Sandra T Sigmon, Lauren Yoshinobu, June Gibb, Charyle Langelle, and Pat Harney. 1991. The will and the ways: development and validation of an individual-differences measure of hope. *Journal of personality and social psychology*, 60(4):570.

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.

Ishan Sanjeev Upadhyay, Anshul Wadhawan, Radhika Mamidi, et al. 2021. Hopeful_Men@ LT-EDI-EACL2021: Hope speech detection using Indic transliteration and transformers. *arXiv preprint arXiv:2102.12082*.

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.

Carolyn M Youssef and Fred Luthans. 2007. Positive organizational behavior in the workplace: The impact of hope, optimism, and resilience. *Journal of management*, 33(5):774–800.