Sentiment Analysis and Homophobia Detection of YouTube Comments

Sushil Ugursandi, Anand Kumar M.

Department of Information Technology, National Institute of Technology Karnataka, Surathkal, 575025

Abstract

Sentiment analysis identifies a graded scale of opinions or emotional responses to a particular subject. Many industries and organisations have been actively researching this area for more than 20 years. The key to understand a user's behaviour while responding on a social media site is to understand their feelings. In contemporary research, a sentence's content is evaluated, the emotion predicted, that helps researchers gain an insight on the reaction of an individual towards a social media topic. Here, a sentence's text data is analysed using several Natural Language Processing techniques before being utilised to categorise this multi-class issue. The detection of homophobia and transphobia in comments on YouTube or other social media sites is second objective of this work. Anger, discomfort, or suspicion against Lesbian, Gay, Bisexual and Transgender people is known as homophobia. It can incite individuals to feel panic, dislike, disrespect, aggression, or wrath. By identifying such occurrences on social media, we can better understand how society works and how people behave. The goal of this work is to analyze social media texts such as comments from YouTube and detect homophobic sentiments using deep learning or machine learning models. In this work 6-layer classification model is used, the F1-Score for sentiment identification using the proposed model in this study was 0.5 on multi-class classification and 0.97 on homophobic/transphobic classification and achieved 1st rank on Homophobic detection in Malayalam language and 4th rank for sentiment analysis in Kannada language.

Keywords

Deep Learning, Homophobic/Transphobic Detection, Neural Networks, Sentiment Analysis, YouTube comments Analysis

1. Introduction

Sentiment analysis uses views acquired from people to determine emotions. With the increase in the reach and availability of the Internet, people increasingly express their thoughts on social media sites like Twitter, Facebook, and YouTube. There is a need to comprehend people's perspectives due to an increase in social media data and Internet users and to understand the emotion within the text, Sentiment Analysis is used. The negative attitude or dislike toward LGBT individuals is referred to as homophobia. LGBT refers to Lesbian, Gay, Bisexual and Transgender people.

Nobody has the right to act aggressively or harshly toward another person. This type of activity or misbehavior can be identified on social media sites using machine learning and deep

FIRE 2022: Forum for Information Retrieval Evaluation, December 9-13, 2022, India

*Corresponding author.

†These authors contributed equally.

sushil.212it032@nitk.edu.in (S. Ugursandi); m_anandkumar@nitk.edu.in (A. K. M.)

learning algorithms. This can be used to flag or prohibit posts that can be dangerous and thus can be used to prevent violence in the community. The practice of violent activities on social media platforms has a negative effect on internet users. Social media is essential for online communication in the digital era because it provides users the flexibility to create, share, and debate anything they choose. Understanding and filtering the content people express on social media is crucial as online communication has grown across languages used across the world.

Main goal of this work is to classify YouTube comments into good, negative, neutral, or mixed categories based on their message polarity on language belongs combinations, especially Malayalam-English, Tamil-English, and Kannada-English. The second objective of this work is to detect homophobic/transphobic nature of comments from social media. The collection includes properly identified homophobic and transphobic language. The training and validation (dev) datasets are in the English, Tamil and Malayalam languages. This work intends to support further research into the detection of homophobic and transphobic content in postings on social media that are published in Tamil, Kannada and Malayalam languages. The dataset for the training, development and testing is obtained from DravidianLangTech 2022.

This competition was organized by DravidianLangTech 2022 [1] and consisted of the following shared task :

- **Task A:** This is a message-level polarity classification task. Given a Youtube comment, systems have to classify it into positive, negative, neutral, or mixed emotions. The participants will be provided development, training, and test dataset code-mixed text in Dravidian languages (Tamil-English, Malayalam-English, and Kannada-English)
- **Task B:** In this shared task, participants will be provided with comments extracted from social media platforms and are expected to develop and submit systems to predict whether it is homophobic/transphobic in nature. The seed data for this task is the Homophobia/Transphobia Detection dataset, a collection of comments from YouTube. This dataset consists of manually annotated comments indicating whether if the text is homophobic/transphobic or not. The participants will be provided development, training, and test dataset in English, Malayalam, and Tamil.

2. Related Work

The below literature survey conducted in the area of Sentiment Analysis and Homophobic detection. People express their ideas on a range of subjects via social media such as YouTube, Facebook, and Twitter according to Jelodar et al. In their work, Jelodar et al. [2] performed Sentiment analysis of YouTube comments on Oscar-nominated movie trailers. Video-sharing platforms like YouTube encourage audience interaction by enabling viewers to rate movie trailers [3]. They organised the movie trailer comments on YouTube in this article. Using these comments, they created three different Sentiment Indexes that gauge the sentiment of movie reviews. Then, projections for the movie's box office take were produced using this sentiment index. Distributors and producers can anticipate audience reaction to a film by examining comments on the trailers. This also leads to a forecast of the profits on the day of release by taking into account existing attitudes. In this work, the collection of movies was predicted

using a unique sentiment analysis technique and sentiment index. Latent topic detection and fuzzy lattice reasoning were used.

G. Prasad et al. [4] analysed the sentiment of cryptocurrency related comments on YouTube. The architecture employed in this study is an ensemble model with stacked KNN, Decision Tree, XG Boost, and Random Forest Classifier components. The meta/base classifier used was logistic regression. T. Mehta et al.[5] used Support Vector Machine, Decision Tree, Linear Regression, Artificial Neural Networks, and Random Forest, predicts the feelings of YouTube Ad View. From a massive collection of comments many of which offer helpful information that raises the posted content's rating levels [6]. 2022 Vinit kumar et al. [7] worked to evaluate significant elements of YouTube videos for the Koha and DSpace apps. Information searchers may now substantially benefit from this since YouTube allows content providers the chance to share their expertise and experiences via their work. These days, individuals routinely search YouTube for answers to their queries or tutorials that might help them better comprehend certain concepts. Analyzing the reactions of the viewers to the videos was another aim of their study.

Homophobia detection involves detecting whether homophobic or not [8]. N. Ashraf et al. [9] involved the use of TF/IDF along with a bigram model. This helped in the vectorization of comments. After this, Support Vector Machines were used. D. Nozza et al. [10] made use of ensemble modeling and data augmentation for high class imbalance. Fine-tuning was used on BERT, RoBERTa and HateBERT. A weighted majority vote was used on the predictions done. A popular topic of discussion is the private American company's launch of the Starlink satellite in [11] by A.M. Putri et al. The satellite launch footage was disseminated via the YouTube channel. Many people commented on the video of the satellite launch, and the comments section on the video had a wide range of opinions. Therefore, this study employed deep learning-based sentiment analysis to look at how internet users responded to the Starlink satellite launch in their YouTube comments. This study comprised 22,000 YouTube comments in total. It uses the Long Short Term Memory model (LSTM). This study produced an accuracy value for the LSTM model utilising various activation and optimization techniques. This study [11] found that the maximum accuracy was 86% when using the LSTM model, Softmax activation function, and Adam's optimization.

H. Bhuiyan et al. [12] claims that YouTube is one of the most well-known social networking sites where users can post, like, comment on, and watch videos. This ranking often preserves the popularity, applicability, and quality of the video. Unrelated or subpar films frequently appear higher in search results due to the quantity of views or likes, which is absurd. To address this issue, they provided a sentiment analysis technique based on Natural Language Processing (NLP) for user comments. This method assisted in identifying the most relevant and popular YouTube video for the specified search. An analysis of the performance of the proposed method in terms of its precision in identifying relevant, popular, and high-quality videos. R. F. Alhujaili at al. [6] identified the sentiments by machine learning and Natural Language Processing algorithms. Numerous academic endeavours using two classes—positive or negative, three classes—two with neutral , or multiple classes have been made (happy, sad, fear, surprise, and anger). However, picking the most precise model might be challenging. As a result, efforts have been undertaken to use sentiment analysis of YouTube comments to determine the polarity. This work examines the methods and techniques for sentiment analysis that may be used on YouTube videos. It also lists and categorises a variety of strategies that are

beneficial for sentiment analysis and data mining investigations.

3. Methodology

3.1. Dataset

Data is obtained from DravidianLangTech 2022 [13] for the following shared tasks[1].

- **Task A:** To find positive, negative, neutral or mixed emotion for the given YouTube comment in Dravidian language.
- **Task B:** To predict homophobic/transophobic nature of comments extracted from social media.

Code-Mixed Dravidian languages used are Kannada-English, Tamil-English and Malayalam-English. The Kannada-English sample dataset is displayed in the Fig 1 along with their class label.

text	category
ಒಂದು ದೇಶದ ಮುಂದುವರಿಯುವುದು ಅದರ ಆರ್ಥಿಕ ಸ್ಥಿತಿಯನ್ನ...	Negative
ಕನ್ನಡದಲ್ಲಿ ಡೈಲಿ ಟಿಕ್ ಅಪ್ಡೇಟ್ಸ್ಪಡೆಯಲು ಸಬ್ಸ್ಕ್ರ...	Positive
Super sar song	not-Kannada
Tiktokers present situation... ಗನೋಡುವವರು ಯಾರು ...	Negative
Super ಸಾಂಗ್ ವೆರಿ ನೈಸ್....	Positive
...	...
@A.R.W tumbad tanhaji andhadhun aise bahot h...	not-Kannada
പൊളി ഡാൻസ് രക്ഷിത് ഷെട്ടി മാസ്സ്	not-Kannada
Bro...nNeen este Roast madudru...China ne beku...	Negative

Figure 1: Kannada-English dataset for Sentiment Analysis

Fig 2 describes the training dataset in Tamil-English language for sentiment analysis.

Similarly the another dataset sample in Malayalam-English language consists of 3 classes 'Non-anti-LGBT+ content', 'Homophobic' and 'Transphobic'.

In Figures 3 and 4 are the training dataset for the second objective of this work which is Homophobic/Transphobic detection on multilingual social media comments.

The overall count for each class label for sentiment analysis and homophobic detection, respectively, is described in Table 1 and Table 2 for all languages.

text	category
Vani bhojam fans hit like solli 500 like Vangi...	unknown_state
I love you ajith very I like	Positive
ennaya trailer Ku mudi Ellam nikkudhu... Vera ...	Positive
Vijay Annaa Ur Maasssssss Therrrrriiiii	Positive
நம்ப நடே நாசாமா தான் போச்சி	Negative
...	...
ஒருவர் செய்த தவறுக்காக எல்லாரையும் பழி போடுவது...	Mixed_feelings
"Repeated Mode to Watching Theri Trailer & The...	Positive
yeevanikuachu pula kutti erudhal oodi poyidu....	Negative
She is looking like laughing budha	Positive
NORTH OR SOUTH KGF IS THE BEST!!	Positive

Figure 2: Tamil-English dataset for Sentiment Analysis

Table 1
Count of instance for each class present in training dataset for Sentiment Analysis

Kannada-English		Tamil-English		Malayalam-English	
Class	Count	Class	Count	Class	Count
Unknown_state	711	Unknown_state	5628	Unknown_state	5279
Positive	2823	Positive	20070	Positive	6421
Negative	1188	Negative	4271	Negative	2105
Mixed_feelings	1722	Mixed_feelings	4020	Mixed_feelings	926
Not_Kannada	916	Not_Tamil	1667	Not-Malayalam	1157

Table 2
Count of instance for each class present in training dataset for Homophobic/Transphobic detection

Tamil-English		Malayalam-English	
Class	Count	Class	Count
Non-anti-LGBT+ content	3438	Non-anti-LGBT+ content	692
Homophobic	311	Homophobic	133
Transphobic	112	Transphobic	41

4. Proposed Model

The algorithm used to generate the categorization report is is illustrated by Fig 5. More details about the precise functioning of the model are given in the subsequent sections of this work.

text	category
@Kadhayalla Nijam thaa gay Punda Apo ne confi...	Homophobic
love you too Nanba	Non-anti-LGBT+ content
Semma super 👍👍👍👍	Non-anti-LGBT+ content
Ava ka elom kadavuluka	Non-anti-LGBT+ content
Bro gay persons kalyanam panninaa kozhanda pet...	Non-anti-LGBT+ content
...	...
Antha akka romba caring	Non-anti-LGBT+ content
@Dan_Lei Nii Pooi Vidurayaa Unakku Yaarum v...	Non-anti-LGBT+ content
Frnds neenga antha yedathula a irrunthaalum am...	Non-anti-LGBT+ content
Looking pretty 😍	Non-anti-LGBT+ content
Apadi kai vekkumbothu kevalama yevono thittita...	Non-anti-LGBT+ content

Figure 3: Tamil-English dataset for Homophobic/Transphobic detection

text	category
ഇവരും ഇവരുടെ സംഗപരിവാരങ്ങളും ചേർന്ന് ഈ നാട് നാ...	Homophobic
കഷ്ടം ഇതിനൊന്നും കോടതി അനുമതി നൽകാൻ പാടില്ല ഉണ...	Homophobic
ഏതോ പുരാണ കാലം മുതൽ ഇന്ത്യയിലും അറേബ്യയിലും ചൈ...	Non-anti-LGBT+ content
Deepak Jayadevan schoolile rss adyapaganmaran ...	Non-anti-LGBT+ content
muzhuvan bahumanathode ningal ellavareyum resp...	Non-anti-LGBT+ content
മരിക്കുക ഇവനെ ഒകെയ് അടിച്ചു കൊല്ലാൻ ആ നാട്ടിൽ ...	Homophobic
Rose chechi pwoli	Non-anti-LGBT+ content
Lgbk nere rape epozhum nadakunond avark swatha...	Non-anti-LGBT+ content
അവസാനിക്കാത്തതായി ഒന്നുമില്ല സ്വയം രക്ഷപ്പെടാൻ...	Non-anti-LGBT+ content

Figure 4: Malayalam-English dataset for Homophobic/Transphobic detection

4.1. Pre-processing

The text data is not readable by system so it needs to be converted into numeric data to make it understandable by system and classify accordingly. The model which is used for classification of sentiments and homophobic detection requires numerical value. Tokenizer is a Python package that turns all text into discrete integer values after cleaning the text by removing stopwords and

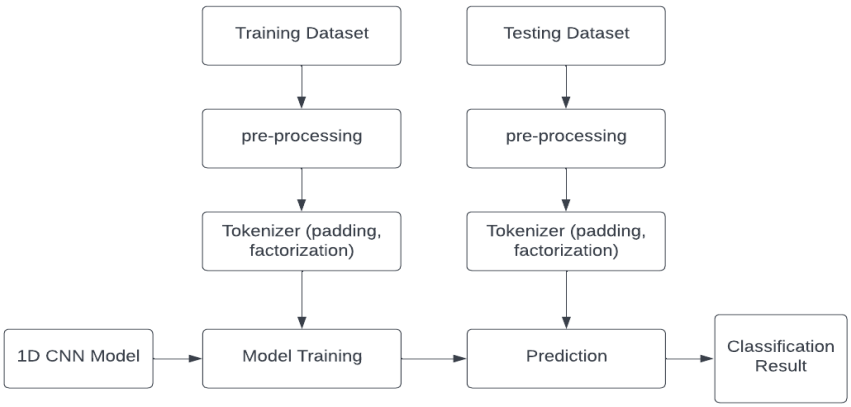

Figure 5: Proposed Model

symbols. This is done as part of the dataset's pre-processing. To fit on the specified input size, post-tokenization padding is used to lengthen each sentence to the same number of characters. The class is one hot encoded in the same manner, creating distinct numerical values according to the associated class label.

Tokenizer library from keras framework is used in this work to tokenize the words using the dictionary size 20000. Padding is applied to make all sentence of same length. Post padding is used with padding size of 94 which fills multiple zeros at end to until it's length reaches to required number.

4.2. Layers

A 6-layer deep learning model is used for this classification problem as shown in Fig 6 The first layer is embedding layer it is used to translate categorical information into integers, we employ one-hot encoding. In order to do this, It generates sample features for each category and fill them with 0s and 1s which generates a vector of results, this layer generates a new layer input over a particular geographic dimension, this process is done by convolutional layer. There is a kernel of fixed size which computes the vector data on that region. After computing the vector data, a fixed size kernel of one dimension iterates over the entire vector and creates a new layer. The third layer is a Max Pooling layer which is used to calculate the maximum value in a vector from each feature value in a matrix. Further the two dimensional matrix data are subsequently transformed into one dimensional values using a flatten layer. The Dropout layer is just a filter which leaves all other neurons unaltered while eliminating particular neuron contributions to the subsequent layer.

4.2.1. Embedding layer

The Embedding layer looks for the embedding vector for each word-index using the vocabulary that has been integer-encoded. The Embedding layer is used from keras framework with the

embedding dimension of 64 and sequence length of 270.

4.2.2. Hyperparameters used

For training of the model the batch size used is 64, using Nadam optimizer and the loss function from keras library CategoricalCrossentropy is applied. Total 50 epochs were used to train this model.

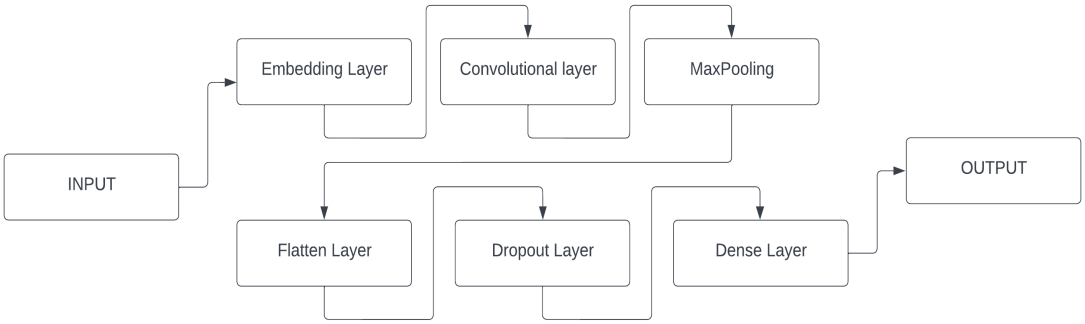

Figure 6: 6-Layer Classification Model

5. Results and Discussion

In this section the results of experiment performed for ”Sentiment analysis and homophobic/-transphobic detection in YouTube comments” is shown.

5.1. Results on validation dataset

Metrics on the prediction from model (Precision, Recall and F-1 Score) is used to display the classification report on validation dataset. Table 3, Table 4 and Table 5 shows the results of Sentiment Analysis on validation dataset for Kannada-English, Tamil-English and Malayalam-English respectively.

These are the results obtained on the validation dataset shown on Table 5 on Malayalam-English language for Sentiment analysis.

For the task of Homophobic detection, the classification results on validation dataset for Tamil language are shown in Table 6 and Malayalam language is shown in Table 7.

Table 3
Result on Sentiment Analysis - (Kannada-English)

Class	Precision	Recall	F1-Score
Unknown_state	0.38	0.48	0.42
Positive	0.69	0.71	0.70
Negative	0.62	0.45	0.52
Mixed_feelings	0.21	0.17	0.19
Not_Kannada	0.54	0.63	0.58
macro avg	0.49	0.49	0.48
weighted avg	0.58	0.58	0.58
Accuracy	–	–	0.58

Table 4
Result on Sentiment Analysis - (Tamil-English)

Class	Precision	Recall	F1-Score
Unknown_state	0.36	0.41	0.38
Positive	0.71	0.75	0.73
Negative	0.40	0.37	0.38
Mixed_feelings	0.25	0.16	0.20
Not_Tamil	0.25	0.16	0.20
macro avg	0.46	0.43	0.44
weighted avg	0.56	0.57	0.57
Accuracy	–	–	0.57

Table 5
Result on Sentiment Analysis - (Malayalam-English)

Class	Precision	Recall	F1-Score
Unknown_state	0.71	0.65	0.68
Positive	0.71	0.77	0.74
Negative	0.60	0.55	0.57
Mixed_feelings	0.40	0.41	0.41
Not_Malayalam	0.73	0.72	0.73
macro avg	0.63	0.62	0.63
weighted avg	0.68	0.68	0.68
Accuracy	–	–	0.68

5.2. Result by Organizers

Among all these experiment on different languages, three of the systems experimented on were submitted. The result obtained from them is shown in Table 8. Using this model 4th rank is achieved for first task on Kannada-English language. For the same problem on Malayalam-English language, the output of this model got 5th rank. For second task, Homophobic/Transphobic detection, 1st Rank is achieved using this model on Malayalam language.

Table 6
Result on Homophobic detection - (Tamil-English)

Class	Precision	Recall	F1-Score
Homophobic	0.34	0.27	0.30
Transphobic	0.62	0.26	0.37
Non-anti-LGBT+content	0.92	0.96	0.94
macro avg	0.63	0.50	0.54
weighted avg	0.87	0.89	0.87
Accuracy	–	–	0.89

Table 7
Result on Homophobic detection - (Malayalam-English)

Class	Precision	Recall	F1-Score
Homophobic	0.97	0.95	0.96
Transphobic	0.99	0.99	0.99
Non-anti-LGBT+content	0.98	1.00	0.99
macro avg	0.98	0.98	0.98
weighted avg	0.99	0.99	0.99
Accuracy	–	–	0.99

Table 8
Results published by organizers

Objective	Precision	Recall	F1-Score	Rank
Homophobic Detection (Malayalam-English)	–	–	0.9744	1
Sentiment Analysis (Kannada-English)	0.48	0.5	0.48	4
Sentiment Analysis (Malayalam-English)	0.5	0.5	0.5	5

5.3. Confusion Matrix

For measurement of performance of the model the confusion matrix is shown in Table 9 for Task A in Malayalam language and in Table 10 for Task B in Malayalam language.

Table 9
Confusion Matrix for Sentiment Analysis in Malayalam language

		Predicted				
		Unknown_state	Positive	Negative	Mixed_feelings	Not-Malayalam
Actual	Unknown_state	396	122	37	14	17
	Positive	105	531	29	25	16
	Negative	49	41	127	16	4
	Mixed_feelings	21	28	12	41	0
	Not-Malayalam	16	22	4	1	98

Table 10
Confusion Matrix for Homophobic detection in Malayalam language

		Predicted		
		Homophobic	Non-anti LGBT	Transphobic
Actual	Homophobic	126	7	0
	Non-anti LGBT	3	685	4
	Transphobic	0	0	41

6. Conclusion and Future work

Sentiment Analysis and Homophobic/Transphobic detection were performed. The proposed model with 6 layers was used in this work. The results obtained were calculated using metrics like Precision, Recall and F-1 score. Sentiment analysis was performed on Malayalam-English, Tamil-English and Kannada-English datasets. The Homophobic/Transphobic detection was performed for the languages Tamil-English and Malayalam-English. The accuracy obtained was 58%, 68% and 57% for Sentiment analysis task in Kannada, Malayalam and Tamil respectively and 89% and 99% for Homophobic detection on Tamil and Malayalam language respectively. The proposed one dimensional model predicted good result for both of the Sentiment Analysis and Homophobic detection. Homophobic detection had 3 classes whereas first objective of Sentiment analysis had 5 class-labels and model predicted better result on Homophobic detection. Users from various region express their opinion in their regional language, the future focus of this research will be based on the intermixed language on different regions.

References

[1] K. Shumugavadivel, M. Subramanian, P. K. Kumaresan, B. R. Chakravarthi, B. B, S. Chinnaudayar Navaneethakrishnan, L. S.K, T. Mandl, R. Ponnusamy, V. Palanikumar, M. Balaji J, Overview of the Shared Task on Sentiment Analysis and Homophobia Detection of YouTube Comments in Code-Mixed Dravidian Languages, in: Working Notes of FIRE 2022 - Forum for Information Retrieval Evaluation, CEUR, 2022.

[2] H. Jelodar, Y. Wang, M. Rabbani, S. B. B. Ahmadi, L. Boukela, R. Zhao, R. S. A. Larik, A nlp framework based on meaningful latent-topic detection and sentiment analysis via fuzzy lattice reasoning on youtube comments, Multim. Tools Appl. 80 (2021) 4155–4181.

[3] H. Timani, P. Shah, M. Joshi, Predicting success of a movie from youtube trailer comments using sentiment analysis, in: 2019 6th International Conference on Computing for Sustainable Global Development (INDIACom), 2019, pp. 584–586.

[4] Divesh, G. Prasad, G. Sharma, D. K. Vishwakarma, Sentiment analysis on cryptocurrency using youtube comments, in: 2022 6th International Conference on Computing Methodologies and Communication (ICCMC), 2022, pp. 730–733. doi:10.1109/ICCMC53470.2022.9753723.

[5] T. Mehta, G. Deshmukh, YouTube ad view sentiment analysis using deep learning and machine learning, International Journal of Computer Applications 184 (2022) 10–14. URL: https://doi.org/10.5120%2Fijca2022922078. doi:10.5120/ijca2022922078.

[6] R. F. Alhujaili, W. M. Yafooz, Sentiment analysis for youtube videos with user comments: Review, in: 2021 International Conference on Artificial Intelligence and Smart Systems (ICAIS), 2021, pp. 814–820. doi:10.1109/ICAIS50930.2021.9396049.

[7] M. Deori, V. Kumar, M. K. Verma, Analysis of youtube video contents on koha and dspace, and sentiment analysis of viewers' comments, Library Hi Tech (2021).

[8] B. R. Chakravarthi, R. Priyadharshini, R. Ponnusamy, P. K. Kumaresan, K. Sampath, D. Thenmozhi, S. Thangasamy, R. Nallathambi, J. P. McCrae, Dataset for identification of homophobia and transophobia in multilingual youtube comments, 2021. URL: https://arxiv.org/abs/2109.00227. doi:10.48550/ARXIV.2109.00227.

[9] N. Ashraf, M. Taha, A. Abd Elfattah, H. Nayel, NAYEL @LT-EDI-ACL2022: Homophobia/-transphobia detection for equality, diversity, and inclusion using SVM, in: Proceedings of the Second Workshop on Language Technology for Equality, Diversity and Inclusion, Association for Computational Linguistics, Dublin, Ireland, 2022, pp. 287–290. URL: https://aclanthology.org/2022.ltedi-1.42. doi:10.18653/v1/2022.ltedi-1.42.

[10] D. Nozza, Nozza@LT-EDI-ACL2022: Ensemble modeling for homophobia and transphobia detection, in: Proceedings of the Second Workshop on Language Technology for Equality, Diversity and Inclusion, Association for Computational Linguistics, Dublin, Ireland, 2022, pp. 258–264. URL: https://aclanthology.org/2022.ltedi-1.37. doi:10.18653/v1/2022.ltedi-1.37.

[11] A. M. Putri, D. A. P. Basya, M. T. Ardiyanto, I. Sarathan, Sentiment analysis of youtube video comments with the topic of starlink mission using long short term memory, in: 2021 International Conference on Artificial Intelligence and Big Data Analytics, 2021, pp. 28–32. doi:10.1109/ICAIBDA53487.2021.9689754.

[12] H. Bhuiyan, J. Ara, R. Bardhan, M. R. Islam, Retrieving youtube video by sentiment analysis on user comment, in: 2017 IEEE International Conference on Signal and Image Processing Applications (ICSIPA), 2017, pp. 474–478. doi:10.1109/ICSIPA.2017.8120658.

[13] B. R. Chakravarthi, R. Priyadharshini, R. Ponnusamy, P. K. Kumaresan, K. Sampath, D. Thenmozhi, S. Thangasamy, R. Nallathambi, J. P. McCrae, Dataset for identification of homophobia and transophobia in multilingual youtube comments, arXiv preprint arXiv:2109.00227 (2021).