# Sentiment Analysis For Bengali Using Transformer Based Models

**Anirban Bhowmick**
Universität Hamburg
anirbanbhowmick88@gmail.com

**Abhik Jana**
Universität Hamburg
abhikjana1@gmail.com

## Abstract

Sentiment analysis is one of the key Natural Language Processing (NLP) tasks that has been attempted by researchers extensively for resource-rich languages like English. But for low resource languages like Bengali very few attempts have been made due to various reasons including lack of corpora to train machine learning models or lack of gold standard datasets for evaluation. However, with the emergence of transformer models pre-trained in several languages, researchers are showing interest to investigate the applicability of these models in several NLP tasks, especially for low resource languages. In this paper, we investigate the usefulness of two pre-trained transformers models namely multilingual BERT and XLM-RoBERTa (with fine-tuning) for sentiment analysis for the Bengali Language. We use three datasets for the Bengali language for evaluation and produce promising performance, even reaching a maximum of 95% accuracy for a two-class sentiment classification task. We believe, this work can serve as a good benchmark as far as sentiment analysis for the Bengali language is concerned.

## 1 Introduction

In this era of the World Wide Web, sharing of information, knowledge, opinion, etc. has been increased by a huge margin since the last decade. Internet users are coming forward to review stuff like books, movies, videos, e-commerce products, etc., and are sharing their experiences which in turn help the next in line users to get feedback upfront. This genre of texts brings in the essence of sentiment analysis task which helps in polarity classification, i.e. determining whether a given text expresses positive, negative, or neutral sentiment. Sentiment analysis of texts can be useful for different applications, like detecting cyberbullying (Saravanaraj et al., 2016), hate speech detection (von Boguszewski et al., 2021; Mathew et al., 2021), e-commerce recommendation system (Hwangbo et al., 2018), etc. There has been a substantial amount of work done by the researchers to tackle sentiment analysis for resource-rich languages like English (Pak and Paroubek, 2010; Feldman, 2013), but for low resource languages, such attempts are scarce (Islam et al., 2020; Sazzed, 2020; Siripragrada et al., 2020). In recent times, for low resource languages like Hindi, Telegu, Bengali, Assamese, Manipuri, Indonesian, etc. (Akhtar et al., 2016; Mukku and Mamidi, 2017; Sazzed, 2020; Le et al., 2016; Kumar and Albuquerque, 2021; Meetei et al., 2021; Das and Singh, 2021; Singh et al., 2021; Kumari et al., 2021) and even for English-Hindi, English- Bengali code-mixed languages (Jamatia et al., 2020), researchers have come up with a solution for sentiment analysis tasks. In another work, R et al. (2012) performed cross-lingual sentiment analysis task where the opinion polarity of a text in a language is predicted using classifier trained in another language. The authors report results on two widely spoken Indian languages, Hindi and Marathi. Gupta et al. (2021) used an LSTM-RNN based approach to determine the sentiment of Hindi tweets and also compared their approach with CNN, machine learning, and Lexicon based approaches. Gupta et al. (2021) uses Hindi SentiWordNet (HSWN) proposed by Joshi et al. (2010) as a lexicon generating tool for hindi text. So Hindi being a major Indian language has been explored whereas more insights are still needed in Bengali. In one of the very recent works, Islam et al. (2020) prepare a two-class and a three-class sentiment analysis dataset in Bengali and report performances of multilingual BERT (Devlin et al., 2019) which is impressive. Moving forward in a similar direction, in this paper we apply two pre-trained transformers models namely multilingual



*Proceedings of the 18th International Conference on Natural Language Processing*, pages 481–486
Silchar, India. December 16 - 19, 2021. 

BERT and XLM-Roberta (Conneau et al., 2020) after fine-tuning and conducting the analysis. In addition to the datasets proposed by Islam et al. (2020), we use two other datasets proposed by Sazzed (2020) and Hossain et al. (2021) for our study. We observe that, by applying fine-tuned multilingual BERT and XLM-RoBERTa (for convenience we will refer XLM-Roberta as XLM-R in our paper), we achieve an accuracy of 63%-94% and 68%-95%, while evaluating against these three target datasets leading to state-of-the-art performances. To the best of our knowledge, this is the first attempt at such a comprehensive study for the Bengali language where pre-trained transformer models' applicability (with fine-tuning) has been investigated for sentiment analysis tasks and evaluated against three datasets. All the codes and datasets are made publicly available[1].

## 2 Dataset

For this study, we use three datasets. The details of these datasets are described below.

**Prothom Alo:** This is the first dataset[2] used in this study which is a publicly available dataset created from user comments on 10 popular news topics from an online Bengali news portal, Prothom Alo[3]. This dataset is introduced by Islam et al. (2020), for convenience, we refer to this dataset as 'Prothom Alo'. The authors scrape user comments from news threads and clean to obtain a total of 17,852 user comments. Each of the comments is tagged by Bengali domain experts into one of the following three classes: positive, negative, and neutral. The authors prepare a variant of this dataset as well which has only two classes by removing the neutral class entries. This step results in a dataset for two-class classification with 13,120 entries.

**YouTube-B:** This is a collection of reviews manually annotated from YouTube Bengali drama[4] consisting of 8500 positive reviews and 3307 negative reviews and is introduced by Sazzed (2020). This dataset is a two-class dataset having only positive and negative as labels. We refer to this dataset as 'YouTube-B' for the rest of the paper. 'B' stands for Bengali language.

[1] https://github.com/Anirbanbhk88/BengaliSentimentWithTransformers

[2] https://github.com/KhondokerIslam/Bengali_Sentiment

[3] https://www.prothomalo.com/

[4] https://data.mendeley.com/datasets/p6zc7krs37/4

| #(Classes) | Dataset | Neu | Pos | Neg |
|---|---|---|---|---|
| Three | Prothom Alo | 4732 | 4769 | 8351 |
| Two | Prothom Alo | - | 4769 | 8351 |
| | YouTube-B | - | 8500 | 3307 |
| | Book-B | - | 982 | 1018 |

Table 1: Class Distribution of Bengali sentiment analysis datasets. 'Neu', 'Pos' and 'Neg' represent neutral, positive, and negative classes, respectively.

**Book-B:** This is the third dataset introduced by Hossain et al. (2021). It is a collection of Bengali book reviews collected from web resources such as blogs, Facebook, and e-commerce sites. This dataset is also a two-class dataset (having positive and negative classes) with 2000 entries of book reviews. We refer to this dataset as 'Book-B' for the rest of the paper.

The details of these three datasets are shown in Table 1.

## 3 Proposed Approach

In this study, we consider the state-of-the-art multilingual BERT model (Devlin et al., 2019) and XLM-RoBERTa model (Conneau et al., 2020) for the Bengali sentiment analysis task. First, we use them separately in one of the recent architecture proposed by Islam et al. (2020), where authors use Long Short Term Memory (LSTM) (Cho et al., 2014), Convolutional Neural Network (CNN) (Hochreiter and Schmidhuber, 1997) and Gated Recurrent Unit (GRU) (Cho et al., 2014)) on top of the transformer model. Next, we fine-tune the pre-trained BERT and XLM-RoBERTa model using each of the three datasets separately (as depicted in Figure 1) and analyze the performances. In both the direction of exploration, BERT and XLM-RoBERTa are the core transformers for our analysis. Hence, a summary of both the BERT model and the XLM-RoBERTa (XLM-R) model is described below.

**Description of models:** BERT and XLM-R are unsupervised language models pre-trained on a large corpus. They are transformer-based models which have encoder-decoder architecture and use attention mechanisms to generate a contextualized representation of words. BERT uses a multi-layer bidirectional transformer encoder. Its self-attention layer performs self-attention in both directions. There are several variants of BERT. For example, *bert-base* has 12 transformers layers, 110M total parameters while *bert-large* has 24 transformers layers, 340M total parameters. They are useful

to solve the long-range dependencies which is a key problem faced by sequence to sequence models like Recurrent Neural Networks(RNN). For our proposed approach we use *bert-base-multilingual-cased*, which is a *bert-base* model checkpoint trained on multilingual corpus.

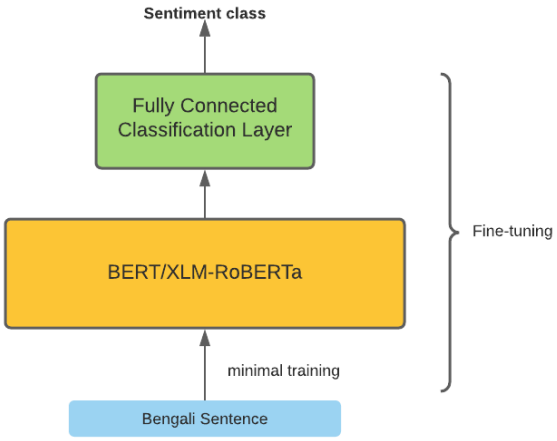


Figure 1: A snapshot of the architecture we fine-tuned for the sentiment analysis task.

| Dataset (#Classes) | Layer on top of BERT | Validation Accuracy |
|---|---|---|
| Prothom Alo (3) | LSTM | 0.58 |
| | CNN | **0.59** |
| | GRU | 0.57 |
| Prothom Alo (2) | LSTM | 0.63 |
| | CNN | **0.74** |
| | GRU | **0.74** |
| YouTube-B (2) | LSTM | 0.85 |
| | CNN | **0.92** |
| | GRU | 0.91 |
| Book-B (2) | LSTM | 0.49 |
| | CNN | **0.91** |
| | GRU | 0.86 |

Table 2: Accuracy of the framework proposed by (Islam et al., 2020) where LSTM, CNN and GRU are used on top of multilingual BERT.

XLM indicates a cross-lingual language model. XLM-RoBERTa (XLM-R) is a pre-trained multilingual model that is considered to be superior over multilingual BERT when evaluated against various NLP tasks. One probable reason could be that XLM-R is trained using a much bigger corpus. XLM-R is also trained in approximately 100 languages. Similarly incase of XLM-R, in our approach we use *xlm-roberta-large* checkpont of pre-trained XLM-R model. We opt for the idea of fine-tuning BERT and XLM-R models especially for low resource language like Bengali, as fine-tuning can be done with a small amount of training data, and the training process is also less time consuming since we are not training all the layers from scratch. Note that, all these transformer-based models used in our study are adopted from HuggingFace.[5].

| Dataset (#Classes) | Layer on top of XLM-R | Validation Accuracy |
|---|---|---|
| Prothom Alo (3) | LSTM | **0.65** |
| | CNN | 0.37 |
| | GRU | 0.63 |
| Prothom Alo (2) | LSTM | 0.64 |
| | CNN | 0.77 |
| | GRU | **0.79** |
| YouTube-B (2) | LSTM | 0.85 |
| | CNN | **0.90** |
| | GRU | **0.90** |
| Book-B (2) | LSTM | 0.60 |
| | CNN | **0.88** |
| | GRU | 0.84 |

Table 3: Accuracy of the variant of the framework proposed by (Islam et al., 2020) where LSTM, CNN, and GRU are used on top of XLM-R.

| Dataset (#Classes) | Models | Validation Accuracy | Test Accuracy |
|---|---|---|---|
| Prothom Alo (3) | BERT | 0.63 | 0.49 |
| | XLM-R | **0.68** | **0.53** |
| Prothom Alo (2) | BERT | 0.77 | 0.69 |
| | XLM-R | **0.81** | **0.73** |
| YouTube-B (2) | BERT | 0.94 | 0.95 |
| | XLM-R | **0.95** | **0.97** |
| Book-B (2) | BERT | **0.91** | **0.91** |
| | XLM-R | **0.91** | 0.87 |

Table 4: Validation (Val) and Test accuracy(Acc) of fine-tuned BERT and XLM-R models all the three datasets respectively. XLM-R here represents XLM-RoBERTa.

## 4 Experimental Setup

For the series of experiments performed, we first adopt the model from the work by Islam et al. (2020) and use it as a baseline model. This benchmark model consists of a multilingual BERT (*bertbase-multilingual-cased*) pre-trained on multiple languages. Three different deep neural network layers: GRU, LSTM, CNN are used as an extra layer on top of BERT separately to produce three separate architectures. We use their code repository and train the baseline models using the same set of hyper-parameters and attempt to replicate the results. Next, we replace BERT with XLM-R in the same architecture. As XLM-R models we use *xlm-roberta-large*. For this set of exper-

[5]https://huggingface.co/transformers

| Samples | Model | #(Classes) | Target | Prediction |
|---|---|---|---|---|
| প্রথম আলোর এই ডিপারমেন্ট খুব কাঁচা রিপোর্টের দিক দিয়ে । This department of Prothom Alo is very raw in terms of reporting. | BERT-Fine | 3 | Neg | Neg |
| খেলাধুলায় ভ্রাতৃত্ববোধ থাকা প্রয়োজন । রেষারেষি নয় । There needs to be a sense of brotherhood in sports. Not a rivalry. | BERT-Fine | 3 | Pos | Pos |
| সবার সাথেই সেম অবস্থা , বুঝতে হবে আমাদের ন্যাচার The situation is same with everyone, we have to understand nature. | BERT-Fine | 3 | Neu | Pos |
| অসাধারন নাটক এমন টা হয়েছিলো আমার সাথে মত । Such an extraordinary drama. The same happened to me. | BERT-Fine | 2 | Pos | Pos |
| প্রথম আলোর এই ডিপারমেন্ট খুব কাঁচা রিপোর্টের দিক দিয়ে । This department of Prothom Alo is very raw in terms of reporting. | BERT-Fine | 2 | Neg | Pos |
| প্রথম আলোর এই ডিপারমেন্ট খুব কাঁচা রিপোর্টের দিক দিয়ে । This department of Prothom Alo is very raw in terms of reporting. | XLM-R-Fine | 3 | Neg | Neg |
| আমরা এমন রেসারেসি চাই না , সবার সাথে বন্ধুত্ব পূর্ণ সম্পর্ক চাই । We do not want such a race, we want a friendly relationship with everyone. | XLM-R-Fine | 3 | Pos | Pos |
| ঘুমন্ত ব্যক্তিকে জাগানো যায় কিন্তু জাগ্রতকে নয় The sleeping person can be awakened but not watchful. | XLM-R-Fine | 3 | Pos | Neu |
| শুধুই ভালোবাসা নিশ ভাই Take Only love brother. | XLM-R-Fine | 2 | Pos | Pos |
| কুরুচিপূরণ শব্দে ভরতি বই A book full of ugly words. | XLM-R-Fine | 2 | Neg | Pos |

Table 5: Sample predictions for fine-tuned BERT and XLM-RoBERTa extracted from different datasets. 'Target' column represents gold standard class as per dataset and 'Prediction' column represents predicted class by our models. 'Neu', 'Pos' and 'Neg' represents the neutral, positive and negative class.

iments, we use a learning rate of $5e^{-}04$.

We also perform another set of experiments where we use pre-trained BERT(*bert-base-multilingual-cased*) and XLM-R(*xlm-roberta-large*) and fine-tune them. For all the fine-tuning experiments a batch size of 16, a learning rate of $2e^{-}05$, and a categorical cross-entropy loss function are used. We use Adam optimizer (Kingma and Ba, 2015) for all the experiments. More details of hyper-parameters used are mentioned in Table 1 of supplementary material.

## 5 Results and Discussion

Even though our primary aim is to investigate the applicability of fine-tuned multilingual BERT and XLM-RoBERTa for Bengali sentiment analysis task, we start our experiment with one of the most recent baseline models proposed by Islam et al. (2020). As Islam et al. (2020) perform all their evaluation on their proposed dataset, Prothom Alo, we first reproduce their result on the same dataset which is presented in the upper half of Table 2. In addition to that, we also evaluate their models on Youtube-B and Book-B datasets as well which are presented in the bottom half of Table 2. We observe BERT with CNN produces an accuracy as high as 0.92 and 0.91 for Youtube-B and Book-B, respectively. Note that, in the study done by Islam et al. (2020), authors report accuracy for the validation set. Therefore, to make a fair comparison we also report the same. Next, we investigate further by using the same model architectures but instead of using multilingual BERT, replacing it with XLM-RoBERTa (XLM-R). The performances of this modified architecture over all three datasets are presented in Table 3. The result shows, that replacing multilingual BERT with XLM-R improves the performance for Prothom Alo dataset (for both three class and two-class classification tasks) by a maximum of 7%. On the other hand, for Youtube-B and Book-B datasets the performance marginally reduces.

Such inconsistencies in performances over different models lead to our next step which deals with fine-tuning multilingual BERT and RoBERTa using three datasets. Note that, in this approach, we do not use any of the LSTM, CNN, and GRU layers on top of the transformer layers as it was done by Islam et al. (2020) as we attempt to show that rather than implementing custom and complex architectures working well on a specific task, simply fine-tuning a transformer is an easier, better alternative. The results of this approach are presented in Table 4. We see that, for 'Prothom Alo' (both two and three classification tasks) fine-tuned XLM-R beats all the previous approaches discussed so far by a significant margin and achieves validation accuracy of 0.68 for the three-class classification task and 0.81 for the two-class classification task. For 'Youtube-B' fine-tuned XLM-R pro-

duces an accuracy of 0.95 whereas for 'Book-B' it produces an accuracy of 0.91 which looks promising.

Note that, to have a fair comparison with the most recent baseline models proposed by (Islam et al., 2020), we report validation accuracy following their performance measures and we see fine-tuned XLM-R outperforms this baseline by a significant margin for the 'Prothom Alo' dataset. In addition, we also report test accuracy in the last column of Table 4. Fine-tuned BERT/XLM-R produces substantially improved performances over closest baselines which will serve as new state-of-the-art performance for these three datasets.

We further investigate a few predicted samples from different datasets to check for the cases that were predicted wrongly by fine-tuned BERT or XLM-R . Few correctly predicted and wrongly predicted samples are presented in Table 5. Even though the overall fine-tuned BERT/XLM-R model performs well, there are certain cases where these models get confused and predict wrongly. In most such cases, the Bengali sentence either contains an ambiguous word or it contains two words from different polarity or it contains some sort of philosophy the meaning of which depends on human interpretation. Taking care of these such cases could be immediate future work.

## 6 Conclusion

In this paper, we conduct an experimental study showing the applicability of multilingual BERT and XLM-R (with fine-tuning) for the Bengali sentiment analysis task. We use three datasets to evaluate the models and obtain promising performances for all three datasets. The immediate future step would be investigating the erroneously classified cases and trying to find the reason behind such errors and mitigate it. Broadly, we plan to investigate sentiment analysis for other low-resource languages like Tamil, Oriya, Gujrati, etc and attempt to propose variants of transformer based models.

## References


Md Shad Akhtar, Ayush Kumar, Asif Ekbal, and Pushpak Bhattacharyya. 2016. A hybrid deep learning architecture for sentiment analysis. In *Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers*, pages 482–493, Osaka, Japan. The COLING 2016 Organizing Committee.

Niklas von Boguszewski, Sana Moin, Anirban Bhowmick, Seid Muhie Yimam, and Chris Biemann. 2021. How hateful are movies? a study and prediction on movie subtitles. In *Proceedings of the 17th Conference on Natural Language Processing (KONVENS 2021)*, pages 37–48, Düsseldorf, Germany. KONVENS 2021 Organizers.

Kyunghyun Cho, Bart van Merriënboer, Caglar Gulcehre, Dzmitry Bahdanau, Fethi Bougares, Holger Schwenk, and Yoshua Bengio. 2014. Learning phrase representations using RNN encoder–decoder for statistical machine translation. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 1724–1734, Doha, Qatar. Association for Computational Linguistics.

Alexis Conneau, Kartikay Khandelwal, Naman Goyal, Vishrav Chaudhary, Guillaume Wenzek, Francisco Guzmán, Edouard Grave, Myle Ott, Luke Zettlemoyer, and Veselin Stoyanov. 2020. Unsupervised cross-lingual representation learning at scale. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 8440–8451, Online. Association for Computational Linguistics.

Ringki Das and Thoudam Doren Singh. 2021. A step towards sentiment analysis of assamese news articles using lexical features. In *Proceedings of the International Conference on Computing and Communication Systems: I3CS 2020, NEHU, Shillong, India*, volume 170, page 15. Springer.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 4171–4186, Minneapolis, MN, USA. Association for Computational Linguistics.

Ronen Feldman. 2013. Techniques and applications for sentiment analysis. *Commun. ACM*, 56(4):82–89.

Vedika Gupta, Nikita Jain, Shubham Shubham, Agam Madan, Ankit Chaudhary, and Qin Xin. 2021. Toward integrated cnn-based sentiment analysis of tweets for scarce-resource language—hindi. *ACM Trans. Asian Low-Resour. Lang. Inf. Process.*, 20(5).

Sepp Hochreiter and Jürgen Schmidhuber. 1997. Long Short-Term Memory. *Neural Comput.*, 9(8):1735—1780.

Eftekhar Hossain, Omar Sharif, and Mohammed Moshiul Hoque. 2021. Sentiment polarity detection on bengali book reviews using multinomial naive bayes. In *Progress in Advanced Computing and Intelligent Engineering*, pages 281–292. Springer.

Hyunwoo Hwangbo, Yang Sok Kim, and Kyung Jin Cha. 2018. Recommendation system development for fashion retail e-commerce. *Electronic Commerce Research and Applications*, 28:94–101.

Khondoker Ittehadul Islam, Md Saiful Islam, and Md Ruhul Amin. 2020. Sentiment analysis in bengali via transfer learning using multi-lingual bert. In *2020 23rd International Conference on Computer and Information Technology (ICCIT)*, pages 1–5. IEEE.

Anupam Jamatia, Steve Durairaj Swamy, Björn Gambäck, Amitava Das, and Swapan Debbarma. 2020. Deep learning based sentiment analysis in a code-mixed english-hindi and english-bengali social media corpus. *International Journal on Artificial Intelligence Tools*, 29(05):2050014.

Aditya Joshi, AR Balamurali, Pushpak Bhattacharyya, et al. 2010. A fall-back strategy for sentiment analysis in hindi: a case study. *Proceedings of the 8th ICON*.

Diederik P. Kingma and Jimmy Ba. 2015. ADAM: A Method for Stochastic Optimization. In *3rd International Conference on Learning Representations, ICLR 2015*, pages 1–15, San Diego, CA, USA.

Akshi Kumar and Victor Hugo C Albuquerque. 2021. Sentiment analysis using xlm-r transformer and zero-shot transfer learning on resource-poor indian language. *Transactions on Asian and Low-Resource Language Information Processing*, 20(5):1–13.

Divya Kumari, Asif Ekbal, Rejwanul Haque, Pushpak Bhattacharyya, and Andy Way. 2021. Reinforced nmt for sentiment and content preservation in low-resource scenario. *Transactions on Asian and Low-Resource Language Information Processing*, 20(4):1–27.

Tuan Anh Le, David Moeljadi, Yasuhide Miura, and Tomoko Ohkuma. 2016. Sentiment analysis for low resource languages: A study on informal indonesian tweets. In *Proceedings of the 12th Workshop on Asian Language Resources (ALR12)*, pages 123–131.

Binny Mathew, Punyajoy Saha, Seid Muhie Yimam, Chris Biemann, Pawan Goyal, and Animesh Mukherjee. 2021. HateXplain: A Benchmark Dataset for Explainable Hate Speech Detection. *Proceedings of the AAAI Conference on Artificial Intelligence*, 35(17):14867–14875.

Loitongbam Sanayai Meetei, Thoudam Doren Singh, Samir Kumar Borgohain, and Sivaji Bandyopadhyay. 2021. Low resource language specific pre-processing and features for sentiment analysis task. *Language Resources and Evaluation*, pages 1–23.

Sandeep Sricharan Mukku and Radhika Mamidi. 2017. ACTSA: Annotated corpus for Telugu sentiment analysis. In *Proceedings of the First Workshop on Building Linguistically Generalizable NLP Systems*, pages 54–58, Copenhagen, Denmark. Association for Computational Linguistics.

Alexander Pak and Patrick Paroubek. 2010. Twitter as a corpus for sentiment analysis and opinion mining. In *LREc*, volume 10, pages 1320–1326.

Balamurali R, Aditya Joshi, and Pushpak Bhattacharyya. 2012. Cross-lingual sentiment analysis for indian languages using linked wordnets. pages 73–82.

A Saravanaraj, JI Sheeba, and S Pradeep Devaneyan. 2016. Automatic detection of cyberbullying from twitter. *International Journal of Computer Science and Information Technology & Security (IJCSITS)*.

Salim Sazzed. 2020. Cross-lingual sentiment classification in low-resource Bengali language. In *Proceedings of the Sixth Workshop on Noisy User-generated Text (W-NUT 2020)*, pages 50–60, Online. Association for Computational Linguistics.

Thoudam Doren Singh, Telem Joyson Singh, Mirinso Shadang, and Surmila Thokchom. 2021. Review comments of manipuri online video: Good, bad or ugly. In *Proceedings of the International Conference on Computing and Communication Systems: I3CS 2020, NEHU, Shillong, India*, volume 170, page 45. Springer.

Shashank Siripragrada, Jerin Philip, Vinay P. Namboodiri, and C V Jawahar. 2020. A multilingual parallel corpora collection effort for Indian languages. In *Proceedings of The 12th Language Resources and Evaluation Conference*, pages 3743–3751, Marseille, France. European Language Resources Association.