IndicFed: A Federated Approach for Sentiment Analysis in Indic Languages

Jash Mehta*, Deep Gandhi*, Naitik Rathod, Sudhir Bagul‡
Dwarkadas J. Sanghvi College of Engineering
Mumbai, India
{jashmehta3300,thisisdeepgandhi,naitikrathod18}@gmail.com
‡sudhir.bagul@djsce.ac.in

Abstract

The task of sentiment analysis has been extensively studied in high-resource languages. Even though sentiment analysis is studied for some resource-constrained languages, the corpora and the datasets available in other low resource languages are scarce and fragmented. This prevents further research of resource-constrained languages and also inhibits model performance for these languages. Privacy concerns may also be raised while aggregating some datasets for training central models. Our work tries to steer the research of sentiment analysis for resource-constrained languages in the direction of Federated Learning. We conduct various experiments to compare server based and federated approaches for 4 Indic Languages - Marathi, Hindi, Bengali, and Telugu. Specifically, we show that a privacy preserving approach, Federated Learning surpasses traditional server trained LSTM model and exhibits comparable performance to other servers-side transformer models.

1 Introduction

With the proliferation of opinionated user data on social media platforms (Murphy et al., 2014), capturing user emotions could help in decision making and determining public opinion on cultural, social, and political agendas (Zhao et al., 2016; Liu, 2012). This has prompted research into sentiment analysis and opinion mining for English (e.g. Thelwall et al., 2010, 2012; Li and Lu, 2017; Hussein, 2018; Li and Lu, 2019; Li et al., 2019; Hoang et al., 2019; Ruz et al., 2020; Chen et al., 2021), which is aided by the availability of large-scale, centralized training datasets. However, there is a pressing need to work on NLP beyond resource-rich languages due to cultural, linguistic, and societal factors (Ruder, 2020).

Sentiment Analysis in low-resource Indic languages has posed a challenge to the research community due to the absence of large-scale centralized datasets. Moreover, to due to increasing concerns and regulations about data privacy (e.g. GDPR (Regulation, 2016)), emerging data has been much more fragmented. It resides in decentralized private silos across different client devices. To abide by such regulations and respect the privacy of users, we must assume that these private data silos can not be shared either with other clients or with the centralized server. Hence, it is exigent to tackle these challenges and study the problem of sentiment analysis in a much more realistic setting - i.e., training models on distributed data silos across different clients to maintain data privacy.

Federated Learning (FL) (McMahan et al., 2017), is a distributed learning paradigm which aims to enable individual clients to train their models collaboratively while keeping their local data private. Instead of accumulating data on a centralized server for training the model, each client sends its model parameters to the server, which updates and sends back the global model to all clients in each round. Since the raw data always remains on the client device and is never shared, FL offers promising solution to the above challenges, particularly in resource poor languages where collection of large-scale training data is difficult.

Previous works for sentiment analysis have relied on traditional server-based architectures and have been centered around resource rich languages. However, such models risk leakage of highly sensitive user-generated data. Thus, we propose a privacy-preserving approach, Federated Learning for sentiment analysis in 4 Indic languages. To the best of our knowledge, our work is the first effort towards Federated learning on Marathi, Bengali, and Telugu datasets; and also towards sentiment analysis in Hindi.

* indicates equal contribution

Proceedings of the 18th International Conference on Natural Language Processing, pages 487–492
Silchar, India. December 16 - 19, 2021. ©2021 NLP Association of India (NLPAI)

Language	Dataset	Reference	# Classes	# Examples			
				Train	Test	Dev	Total
Marathi (mr)	L3CubeMahaSent	Kulkarni et al. (2021)	3	12,114	2,250	1,500	15,864
Telugu (te)	ACTSA	Mukku and Mamidi (2017)	3	3,784	812	812	5,408
Bengali (bn)	ABSA Cricket	Rahman et al. (2018)	3	2,085	380	372	2,837
	ABSA Restaurant	Rahman et al. (2018)	3	1,365	219	224	1,808
	YouTube Comments	Tripto and Ali (2018)	3	1,957	420	419	2,796
	SAIL	Patra et al. (2015)	3	697	204	98	999
	BengFastText	Karim et al. (2020)	2	5,510	1,532	1,378	8,420
	Combined	Hasan et al. (2020)	3	9,901	2,755	2,491	15,147
Hindi (hi)	Hindi Sentiment Analysis	Sinha (2019) [1]	3	6,353	1,362	1,362	9,077

Table 1: Summary of different datasets

We use of publicly available datasets for sentiment analysis in Marathi (Kulkarni et al., 2021), Telugu (Mukku and Mamidi, 2017), Bengali (Hasan et al., 2020), and Hindi [1]. We examine how Federated LSTM model performs, in comparison to 4 server-side centralized models: bi-directional LSTM (Hochreiter and Schmidhuber, 1997), IndicBert (Kakwani et al., 2020b), mBERT (Devlin et al., 2019), and XLM-R (Conneau et al., 2020). We find that the federated learning architecture outperforms the centralized server-side LSTM model and shows comparable performance to the centralized transformer models for all 4 languages under consideration.

The remainder of the paper is organized into prior work (§2), a brief description of the datasets we use (§3), a description of the experimental setup (§4), an in-depth analysis our experiments (§5), and finally a conclusion (§6).

2 Prior Work

Federated Learning: Federated Learning (McMahan et al., 2017) is used for building PPML (Privacy Preserving Machine Learning) models. As proposed by Hard et al. (2018), Federated Learning is useful for preventing bottlenecks when the data is trained on central servers. Some major work is being done for the English language at the intersection of Federated Learning and Natural Language Processing which was also observed in Lin et al. (2021). However, even though benchmarks were established for English language, the task of using it on Indic resource-constrained languages remains relatively unexplored. Recently, Singh et al. (2021) showed that Federated Learning surpasses the baselines for complaint identification in some Indic languages. It is evident from these approaches that using this technique helps achieve better results. Therefore, we consider a total of 4 Indic languages (Hindi, Marathi, Telugu, Bengali) in this paper to conduct Federated Learning.

Sentiment Analysis in Indic Languages: Joshi et al. (2010) talks about various approaches for sentiment analysis in Hindi, and resorts to translating the data to English for sentiment identification due to the issue of constrained resources for Hindi. However, even though Hindi cannot be considered a very resource-constrained language now due to the development of various corpora such as (e.g. Kunchukuttan et al., 2018; Khandelwal et al., 2018; Bafna and Saini, 2021), but a lot of other Indic languages are still resource constrained and corpora for the same are very limited. The datasets for such Indic languages are spread out as observed for Bengali in Table 1. Many approaches have been adopted server trained models for sentiment analysis of the languages being considered here (e.g. Salehin et al., 2020; Regatte et al., 2020; Kulkarni et al., 2021; Jain et al., 2020; Hasan et al., 2020; Kakwani et al., 2020a).

3 Datasets

We use publicly available datasets for sentiment analysis in low-resource Indian languages. Table 1 gives the details of all datasets used in our work. We combine all the Bengali datasets and train our models on the combined dataset.

In case of ACTSA Dataset (Mukku and Mamidi, 2017) and Hindi Sentiment Analysis Dataset [1], we make a stratified split of 70:15:15 to divide the data into train, test, and development sets. For other datasets, we use the original splits which are provided.

[1] https://github.com/sid573/Hindi_Sentiment_Analysis

	mBERT			XLM-R			IndicBERT		
	Acc.	AUC	F1	Acc.	AUC	F1	Acc.	AUC	F1
te	45.81 ± 0.00	49.46 ± 1.66	28.79 ± 0.00	48.45 ± 5.89	54.98 ± 10.34	34.81 ± 13.46	$\mathbf{61.85} \pm \mathbf{0.53}$	$\mathbf{76.61} \pm \mathbf{0.69}$	$\mathbf{61.74} \pm \mathbf{0.62}$
hi	75.57 ± 3.93	87.03 ± 3.83	74.55 ± 5.02	88.39 ± 0.39	94.93 ± 1.38	87.40 ± 2.49	$\mathbf{88.58} \pm \mathbf{1.61}$	$\mathbf{95.73} \pm \mathbf{0.72}$	$\mathbf{88.58} \pm \mathbf{1.62}$
bn	77.54 ± 0.64	80.66 ± 1.01	76.72 ± 0.67	78.30 ± 6.87	81.24 ± 7.58	77.00 ± 8.28	$\mathbf{80.87} \pm \mathbf{0.26}$	$\mathbf{86.27} \pm \mathbf{0.76}$	$\mathbf{80.35} \pm \mathbf{0.23}$
mr	69.98 ± 0.57	83.58 ± 0.43	69.97 ± 0.55	82.47 ± 0.46	92.02 ± 0.62	82.42 ± 0.47	$\mathbf{83.36} \pm \mathbf{0.36}$	$\mathbf{93.51} \pm \mathbf{0.36}$	$\mathbf{83.33} \pm \mathbf{0.35}$

Table 2: Performance of centrally trained models on 5 different seeds. IndicBERT performs better than the others for all languages.

3.1 Pre-processing

To pre-process the data, we lower-case all text and remove numbers, punctuation, and URLS. Since some of the datasets are taken from Twitter, we also remove Twitter specific things like hashtags, @-mentions, and the retweet marker: ”RT:”.

4 Experiments

All the experiments are conducted on Google Colab using a NVIDIA Tesla P100 GPU (16 GB) with 26 GB RAM. The metrics used to compare the results are weighted AUC, weighted F1 and the accuracy score for every model in every variation.

4.1 Central Training

In order to compare results to Federated Learning, we use 4 different models: mBERT (Devlin et al., 2019), XLM-R (Conneau et al., 2020), IndicBERT (Kakwani et al., 2020b), and Bi-LSTM (Hochreiter and Schmidhuber, 1997). We report the mean and standard deviation after training on 5 random seeds. All of these models except the Bi-LSTM architecture are pretrained on the languages being considered in this paper.

Every pretrained model is trained for 25 epochs and the Bi-LSTM model is run for 500 epochs with early stopping. The default learning rate ($lr = 4e - 5$) is used for the pretrained models and for the Bi-LSTM model, the learning rate of 0.01 is set.

4.2 Federated Learning

We use the FedProx algorithm (Li et al., 2018) because it works better in non-iid data where the distribution varies rapidly within the dataset. We conduct various experiments under synthetic-iid (independent and identical distribution) (Li et al., 2018) and non-iid settings. Since, some of these Indic languages cannot be tokenized using general tokenizers such as Spacy[2], we use language specific tokenizers provided by iNLTK (Arora, 2020).

To make the computation cheaper on resource-constrained edge-devices, by the distributed training process, we train it on a basic Bi-LSTM model (Hochreiter and Schmidhuber, 1997) with 2 hidden layers and dropout (Srivastava et al., 2014) set to 0.5. Different client fractions are used to observe the variation of results in the Federated setting too. The client fractions of 10%, 30% and 50% are considered and these clients are always picked randomly for every round.

All the models are run for 500 rounds with early stopping applied on the average training loss. The learning rate is set to 0.01 and the proximal term is set to 0.01 as default (Li et al., 2018).

5 Results

Telugu (te): From Table 4, it can be observed that for the Telugu language, the Federated training process performs much better than the centrally trained LSTM model. The best model chosen for the Federated setting is with $c = 30\%$. Even though the Federated training is trained on 30% of the data every round, the results are better than the central model. Looking at the other models trained for the Telugu language, we also observe from Table 2 that the best performing model IndicBERT (Kakwani et al., 2020b) has comparable results to the Federated LSTM model[3].

Hindi (hi): From Table 4, we find that the federated model performs better on all the metrics than the server-based LSTM model. Even though we achieve better F1 score for $c = 30\%$ (Table 3), we consider the model with $c = 10\%$ as the high performing model for federated learning because of the AUC score and the accuracy. It must be noted that the scores are on the lower side for models trained on non-iid setting in federated learning because every client cluster is intentionally biased to represent one single class unlike the synthetic-iid method.

[2] https://spacy.io/

[3] IndicBERT was pretrained on 674M tokens of Telugu. (Kakwani et al., 2020b)

	c = 10%						**c = 30%**						**c = 50%**					
	IID			non-IID			IID			non-IID			IID			non-IID		
	Acc	AUC	F1	Acc	AUC	F1	Acc	AUC	F1	Acc	AUC	F1	Acc	AUC	F1	Acc	AUC	F1
te	47.62	**76.09**	**55.75**	38.10	64.42	37.74	**66.67**	74.32	54.38	42.86	66.88	25.54	61.91	71.16	49.90	38.10	68.33	32.32
hi	**81.82**	**99.79**	84.18	45.46	92.18	52.14	75.76	99.63	**84.57**	45.45	87.95	49.46	75.76	95.12	76.73	60.61	88.51	62.13
bn	**68.12**	39.38	80.18	53.03	33.15	59.21	65.15	38.65	74.94	46.97	34.16	61.18	65.15	**39.74**	**81.16**	53.03	36.16	64.57
mr	68.52	95.15	79.33	35.19	81.06	38.12	68.52	95.15	79.33	50.00	81.44	51.01	**70.37**	**95.58**	**79.57**	48.15	82.62	62.70

Table 3: Performance of model under federated settings conducted with 3 different client fractions. c is the fraction of clients whose updates are considered in every round. Evidently, for lower dataset sizes, $c = 10\%$ performs comparatively better.

	server-LSTM			**federated-LSTM**		
	Acc	AUC	F1	Acc	AUC	F1
te	51.28	73.91	54.60	66.67	74.32	54.38
hi	78.79	98.92	82.60	81.82	99.79	84.18
bn	60.61	40.92	83.46	68.12	39.38	80.18
mr	65.74	92.03	76.67	70.37	95.58	79.57

Table 4: Comparison between the centrally trained Bi-LSTM model and federated Bi-LSTM model.The federated model is selected based on best results from Table 3

Bengali (bn): Table 4 shows that the Federated Bengali model performs better in terms of accuracy against the centrally trained LSTM but worse in terms of AUC and F1-score. The reason is that the Bengali federated model is trained differently. For the synthetic-IID setting, every client cluster consists data from one specific dataset only and no dataset entries are mixed for every client. Since we use the combined dataset (Hasan et al., 2020), all of the datasets in it have different distributions in terms of categories. We believe that the poor AUC and F1 scores are due to this difference of distribution as the 'neutral' category is absent in some of these member datasets.

Marathi (mr): Looking at Table 4, it is evident that the Federated LSTM performs better on all metrics than the centrally trained LSTM. Since this trend continues across all the languages, we believe that Federated Learning helps learn the data representations better without data bottlenecks and also without sharing any data which might be sensitive.

6 Conclusion

We show that a LSTM model trained using federated learning can outperform an identical server trained LSTM model for 4 Indic languages - Marathi, Bengali, Telugu and Hindi. We also show that federated learning achieves comparable performance to other server trained Transformer architectures[4]. Surprisingly, we find that for smaller datasets, lower client fractions show better performance. To our knowledge, this represents one of the first applications of federated learning in low-resource settings for sentiment analysis. Federated learning offers security and privacy advantages for users by training across a population of highly distributed computing devices while simultaneously improving model performance.

For future work, it would be interesting to train heavier transformer models like IndicBERT, XLM-R, etc. using federated learning which could help to minimize the large gap in accuracy in non-iid settings. Conducting some interpretable evaluation on the intermediate models before updating during Federated training is another important future direction.

References

Gaurav Arora. 2020. iNLTK: Natural language toolkit for indic languages. In *Proceedings of Second Workshop for NLP Open Source Software (NLP-OSS)*, pages 66–71, Online. Association for Computational Linguistics.

Prafulla B Bafna and Jatinderkumar R Saini. 2021. Scaled document clustering and word cloud-based summarization on hindi corpus. In *Progress in Advanced Computing and Intelligent Engineering*, pages 398–408. Springer.

Zhexue Chen, Hong Huang, Bang Liu, Xuanhua Shi, and Hai Jin. 2021. Semantic and syntactic enhanced aspect sentiment triplet extraction. *arXiv preprint arXiv:2106.03315*.

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–

[4]The performance of IndicBERT to the federated LSTM is comparable and the lower results might be because of the difference in architecture complexity.

8451, Online. Association for Computational Linguistics.

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, Minnesota. Association for Computational Linguistics.

Andrew Hard, Kanishka Rao, Rajiv Mathews, Swaroop Ramaswamy, Françoise Beaufays, Sean Augenstein, Hubert Eichner, Chloé Kiddon, and Daniel Ramage. 2018. Federated learning for mobile keyboard prediction. *arXiv preprint arXiv:1811.03604*.

Md Arid Hasan, Jannatul Tajrin, Shammur Absar Chowdhury, and Firoj Alam. 2020. Sentiment classification in bangla textual content: A comparative study. In *2020 23rd International Conference on Computer and Information Technology (ICCIT)*, pages 1–6. IEEE.

Mickel Hoang, Oskar Alija Bihorac, and Jacobo Rouces. 2019. Aspect-based sentiment analysis using BERT. In *Proceedings of the 22nd Nordic Conference on Computational Linguistics*, pages 187–196, Turku, Finland. Linköping University Electronic Press.

Sepp Hochreiter and Jürgen Schmidhuber. 1997. Long short-term memory. *Neural Computation*, 9(8):1735–1780.

Doaa Mohey El-Din Mohamed Hussein. 2018. A survey on sentiment analysis challenges. *Journal of King Saud University-Engineering Sciences*, 30(4):330–338.

Kushal Jain, Adwait Deshpande, Kumar Shridhar, Felix Laumann, and Ayushman Dash. 2020. Indic-transformers: An analysis of transformer language models for indian languages. *arXiv preprint arXiv:2011.02323*.

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*.

Divyanshu Kakwani, Anoop Kunchukuttan, Satish Golla, NC Gokul, Avik Bhattacharyya, Mitesh M Khapra, and Pratyush Kumar. 2020a. inlpsuite: Monolingual corpora, evaluation benchmarks and pre-trained multilingual language models for indian languages. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: Findings*, pages 4948–4961.

Divyanshu Kakwani, Anoop Kunchukuttan, Satish Golla, Gokul N.C., Avik Bhattacharyya, Mitesh M. Khapra, and Pratyush Kumar. 2020b. IndicNLPSuite: Monolingual corpora, evaluation benchmarks and pre-trained multilingual language models for Indian languages. In *Findings of the Association for Computational Linguistics: EMNLP 2020*, pages 4948–4961, Online. Association for Computational Linguistics.

Md Rezaul Karim, Bharathi Raja Chakravarthi, John P McCrae, and Michael Cochez. 2020. Classification benchmarks for under-resourced bengali language based on multichannel convolutional-lstm network. In *2020 IEEE 7th International Conference on Data Science and Advanced Analytics (DSAA)*, pages 390–399. IEEE.

Ankush Khandelwal, Sahil Swami, Syed Sarfaraz Akhtar, and Manish Shrivastava. 2018. Humor detection in english-hindi code-mixed social media content : Corpus and baseline system. *ArXiv*, abs/1806.05513.

Atharva Kulkarni, Meet Mandhane, Manali Likhitkar, Gayatri Kshirsagar, and Raviraj Joshi. 2021. L3CubeMahaSent: A Marathi tweet-based sentiment analysis dataset. In *Proceedings of the Eleventh Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis*, pages 213–220, Online. Association for Computational Linguistics.

Anoop Kunchukuttan, Pratik Mehta, and Pushpak Bhattacharyya. 2018. The IIT Bombay English-Hindi parallel corpus. In *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC-2018)*, Miyazaki, Japan. European Languages Resources Association (ELRA).

Hao Li and Wei Lu. 2017. Learning latent sentiment scopes for entity-level sentiment analysis. In *Thirty-First AAAI Conference on Artificial Intelligence*.

Hao Li and Wei Lu. 2019. Learning explicit and implicit structures for targeted sentiment analysis. *arXiv preprint arXiv:1909.07593*.

Tian Li, Anit Kumar Sahu, Manzil Zaheer, Maziar Sanjabi, Ameet Talwalkar, and Virginia Smith. 2018. Federated optimization in heterogeneous networks. *arXiv preprint arXiv:1812.06127*.

Xin Li, Lidong Bing, Piji Li, and Wai Lam. 2019. A unified model for opinion target extraction and target sentiment prediction. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 33, pages 6714–6721.

Bill Yuchen Lin, Chaoyang He, Zihang Zeng, Hulin Wang, Yufen Huang, Mahdi Soltanolkotabi, Xiang Ren, and Salman Avestimehr. 2021. Fednlp: A research platform for federated learning in natural language processing. *arXiv preprint arXiv:2104.08815*.

Bing Liu. 2012. Sentiment analysis and opinion mining. *Synthesis lectures on human language technologies*, 5(1):1–167.

Brendan McMahan, Eider Moore, Daniel Ramage, Seth Hampson, and Blaise Aguera y Arcas. 2017. Communication-efficient learning of deep networks from decentralized data. In *Artificial intelligence and statistics*, pages 1273–1282. PMLR.

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.

Joe Murphy, Michael W Link, Jennifer Hunter Childs, Casey Langer Tesfaye, Elizabeth Dean, Michael Stern, Josh Pasek, Jon Cohen, Mario Callegaro, and Paul Harwood. 2014. Social media in public opinion research: Executive summary of the aapor task force on emerging technologies in public opinion research. *Public Opinion Quarterly*, 78(4):788–794.

Braja Gopal Patra, Dipankar Das, Amitava Das, and Rajendra Prasath. 2015. Shared task on sentiment analysis in indian languages (sail) tweets-an overview. In *International Conference on Mining Intelligence and Knowledge Exploration*, pages 650–655. Springer.

Md Rahman, Emon Kumar Dey, et al. 2018. Datasets for aspect-based sentiment analysis in bangla and its baseline evaluation. *Data*, 3(2):15.

Yashwanth Reddy Regatte, Rama Rohit Reddy Gangula, and Radhika Mamidi. 2020. Dataset creation and evaluation of aspect based sentiment analysis in Telugu, a low resource language. In *Proceedings of the 12th Language Resources and Evaluation Conference*, pages 5017–5024, Marseille, France. European Language Resources Association.

General Data Protection Regulation. 2016. Regulation eu 2016/679 of the european parliament and of the council of 27 april 2016. Available at: `https://gdpr.eu/`.

Sebastian Ruder. 2020. Why You Should Do NLP Beyond English. `http://ruder.io/nlp-beyond-english`.

Gonzalo A Ruz, Pablo A Henríquez, and Aldo Mascareño. 2020. Sentiment analysis of twitter data during critical events through bayesian networks classifiers. *Future Generation Computer Systems*, 106:92–104.

S. M. Samiul Salehin, Rasel Miah, and Md Saiful Islam. 2020. A comparative sentiment analysis on bengali facebook posts. In *Proceedings of the International Conference on Computing Advancements*, ICCA 2020, New York, NY, USA. Association for Computing Machinery.

Apoorva Singh, Tanmay Sen, Sriparna Saha, and Mohammed Hasanuzzaman. 2021. Federated multi-task learning for complaint identification from social media data. In *Proceedings of the 32nd ACM Conference on Hypertext and Social Media*, HT '21, page 201–210, New York, NY, USA. Association for Computing Machinery.

Nitish Srivastava, Geoffrey Hinton, Alex Krizhevsky, Ilya Sutskever, and Ruslan Salakhutdinov. 2014. Dropout: a simple way to prevent neural networks from overfitting. *The journal of machine learning research*, 15(1):1929–1958.

Mike Thelwall, Kevan Buckley, and Georgios Paltoglou. 2012. Sentiment strength detection for the social web. *Journal of the American Society for Information Science and Technology*, 63(1):163–173.

Mike Thelwall, Kevan Buckley, Georgios Paltoglou, Di Cai, and Arvid Kappas. 2010. Sentiment strength detection in short informal text. *Journal of the American society for information science and technology*, 61(12):2544–2558.

Nafis Irtiza Tripto and Mohammed Eunus Ali. 2018. Detecting multilabel sentiment and emotions from bangla youtube comments. In *2018 International Conference on Bangla Speech and Language Processing (ICBSLP)*, pages 1–6. IEEE.

Jun Zhao, Kang Liu, and Liheng Xu. 2016. Sentiment analysis: mining opinions, sentiments, and emotions.