A Lexicon for Profane and Obscene Text Identification in Bengali

Salim Sazzed
Old Dominion University
Norfolk, VA, USA
ssazz001@odu.edu

Abstract

Bengali is a low-resource language that lacks tools and resources for profane and obscene textual content detection. Until now, no lexicon exists for detecting obscenity in Bengali social media text. This study introduces a Bengali obscene lexicon consisting of over 200 Bengali terms that can be considered filthy, slang, profane or obscene. A semi-automatic methodology is presented for developing the obscene lexicon that leverages an obscene corpus, word embedding, and part-of-speech (POS) taggers. The developed lexicon achieves coverage of around 0.85 for obscene and profane content detection in an evaluation dataset. The experimental results imply that the developed lexicon is effective at identifying obscenity in Bengali social media content.

1 Introduction

The popularity of e-commerce and social media has surged the availability of user-generated content online. Therefore, text analysis tasks such as sentiment classification (Feldman, 2013; Sazzed and Jayarathna, 2019; Yadollahi et al., 2017; Sazzed, 2021b; Sazzed and Jayarathna, 2021), hate speech detection (Poletto et al., 2021; Corazza et al., 2020), profane or abusive content identification (Caselli et al., 2020; Nobata et al., 2016) have received significant attention in recent years. Profanity indicates the usage of taboo or swearing words and is prevalent in social media data across languages (Wang et al., 2014). The presence of swearing, obscene or vulgar words could be linked with hate speech, sexism, and racism. Hence, identifying their presence is important to understanding and monitoring online content. Although the terms profanity, obscenity, swearing, and vulgarity have subtle differences in their meaning, they are closely connected with some overlapping definitions. Thus, in this paper, they have been used interchangeably to refer to filthy content.

A lexicon consisting of a list of words with specific annotations can play an important role in various natural language processing tasks, such as sentiment analysis or inappropriate content identification. A profane or obscene lexicon contains words that convey foul, filthy, and profane meanings (e.g., *ass, bitch*). An obscene lexicon is instrumental for determining profanity, vulgarity, or obscenity in a text. The presence of swearing in English social media has been investigated by various researchers (Wang et al., 2014; Pamungkas et al., 2020). Wang et al. (2014) found that the rate of swear word use in English Twitter is 1.15%, almost double compared to its use in daily conversation (0.5% - 0.7%) as observed in previous work (Jay, 1992). Wang et al. (2014) also reported that in a random sampling, they observed around 7.73% tweets containing swear words. Furthermore, vulgar word identification can help to improve sentiment classification as shown by various researchers (Volkova et al., 2013; Yang and Eisenstein, 2017).

In Bengali, although few works performed abusive content analysis, none of them focused on determining obscenity or generating resources for identifying obscenity. Until now, no lexicon exists in Bengali that can help to identify profanity in text data. Thus, in this work, the goal is to generate resources for obscenity identification.

To construct the Bengali obscene lexicon, we propose a corpus-based semi-automatic approach. From an existing Bengali obscene corpus, utilizing word embedding and POS tagging, the lexicon is created. To demonstrate the efficacy of this lexicon, we categorize a drama review corpus into profane and non-profane categories based on the presence of swear and obscene terms. We observe that the developed lexicon successfully identifies 85.5% of the obscene or profane reviews in the corpus.

Proceedings of Recent Advances in Natural Language Processing, pages 1289–1296
Sep 1–3, 2021.
https://doi.org/10.26615/978-954-452-072-4_145

1.1 Motivation and Contributions

With the rapid growth of user-generated Bengali content on social media and the web, the presence of inappropriate content has become an issue. The content which is not in line with the social norms and expectations of a community needs to be censored. In Bengali, no such resources exist; thus, we focus on building a lexicon consisting of swear or obscene words that can help to identify profane content.

The main contributions of this paper can be summarized as follows-

- We introduce a Bengali obscene lexicon comprised of about 200 swear words. We have made the developed lexicon publicly available for researchers [1].
- We present a semi-automatic methodology for developing a swear lexicon utilizing an obscene corpus and various natural language processing tools.
- We demonstrate that the developed lexicon is effective at profanity detection in Bengali social media content.

2 Related Work

The existence and socio-linguistics characteristics of swearing or cursing in social media have been studied in several studies. Wang et al. (2014) investigated the ubiquity, utility, and contextual dependency of swearing on Twitter. Gauthier et al. (2015) analyzed several sociolinguistic aspects of swearing on Twitter text data. Several studies investigated the relationship between social factors, such as gender with the profanity, and discovered males employ profanity much more often than females (Wang et al., 2014; Selnow, 1985). Other social factors such as age, religiosity, or social status were found to be related to the rate of using vulgar words (McEnery, 2004). Jay and Janschewitz (2008) noticed that the offensiveness of taboo words depends on their context, and found that usages of taboo words in conversational context is less offensive than hostile context. Pinker (2007) classified the use of swear words into five categories: dysphemistic; abusive, using taboo words to abuse or insult someone; idiomatic, using taboo words to arouse the interest of listeners without really referring to the matter; emphatic, to emphasize another word; cathartic, the use of swear words as a response to stress or pain.

Obscenity and profanity filtering has been studied for content filtering, such as parental controls (Weir and Duta, 2012), cyberbullying detectors (Dadvar et al., 2013). A more complex application of obscenity filtering is identifying implicitly abusive content, where both the intention of the author and the usage of obscene language need to be considered (Weir and Duta, 2012).

Research related to the identification of swearing or offensive words has been conducted mainly in English; Therefore, lexicons comprised of offensive words are available in the English language. Pamungkas et al. (2020) created SWAD (Swear Words Abusiveness Dataset), a Twitter English corpus, where abusive swearing is manually annotated at the word level. Their collection consists of 1,511 unique swear words from 1,320 tweets. Razavi et al. (2010) manually collected approximately 2,700 dictionary entries, including phrases and multi-word expressions, which is one of the earliest work offensive lexicon creations. The recent work of lexicon creation for hate speech detection was reported in (Gitari et al., 2015). Another English lexicon of abusive words was provided by (Wiegand et al., 2018).

Eder et al. (2019) explored the vulgar and obscene text in German. They conceived vulgar language is predominantly signaled by an overly lowered language, disgusting and obscene lexicalizations, which is generally banned from any type of civilized discourse. Primarily, it refers to sexual organs and activities, as well as body parts and scatologic expressions.

In Bengali, several works investigated the presence of abusive language in social media data by employing supervised ML classifiers and labeled data (Ishmam and Sharmin, 2019; Banik and Rahman, 2019). Emon et al. (2019) utilized linear support vector classifier (LinearSVC), logistic regression (LR), multinomial naïve Bayes (MNB), random forest (RF), artificial neural network (ANN), recurrent neural network (RNN) with long short term memory (LSTM) to detect multi-type abusive Bengali text. They found RNN outperformed other classifiers by obtaining the highest accuracy of 82.20%.

Chakraborty and Seddiqui (2019) employed ma-

[1] https://github.com/sazzadcsedu/Bangla-Vulgar-Lexicon.git

chine learning and natural language processing techniques to build an automatic system for detecting abusive comments in Bengali. As input, they used Unicode emoticons and Unicode Bengali characters. They applied MNB, SVM, Convolutional Neural Network (CNN) with LSTM and found SVM performed best with 78% accuracy. Sazzed (2020b) created a sentiment lexicon that consists of over 500 Bengali negative opinion words. However, no annotations regarding obscenity or vulgarity were provided for these negative words.

Karim et al. (2020) proposed BengFastText, a word embedding model for Bengali, and incorporated it into a Multichannel Convolutional-LSTM (MConv-LSTM) network for predicting different types of hate speech. They compared BengFastText against the Word2Vec and GloVe embedding by integrating them into several ML classifiers and showed the efficacy of BengFastText for hate speech detection.

Sazzed (2021a) introduced an annotated Bengali corpus of 3000 transliterated Bengali comments categorized into two classes, abusive and non-abusive, 1500 comments for each. For the baseline evaluations, the author employed several supervised machine learning (ML) and deep learning-based classifiers. They observed support vector machine (SVM) shows the highest efficacy for identifying abusive content.

To the best of the author's knowledge, none of these existing works concentrated on recognizing obscene words in Bengali social media content. Besides, no lexicon exists so that profanity or obscenity can be determined without using any annotated data. This work is the first effort to identify profanity in the context of Bengali social media data by introducing a obscene lexicon.

3 Corpora

Two Bengali obscene datasets are used in this study, one for constructing the lexicon (*development corpus*), and the other one is for evaluating the performance of the created lexicon (*evaluation corpus*).

3.1 Development Corpus

The *development corpus* is a subset of a Bengali corpus deposited by (Abu, 2020). This Bengali corpus consists of 10221 user comments which belong to different categories, such as toxic, racism, obscene, insult, etc.

For developing the lexicon, only the obscene comments are used. After discarding the noisy reviews (e.g., empty comments, punctuation only comments, etc.) and reviews that belong to other classes, the *development corpus* consists of 3902 obscene comments (each contains 1-100 words).

Figure 1 represents some examples of the obscene comments from the development corpus.

3.2 Evaluation Corpus

The *evaluation corpus* is a subset of the dataset deposited by Sazzed (2020a). This corpus consists of viewer's comments towards a number of Bengali dramas collected from Youtube. Originally, this dataset contains 8500 positive and 3307 negative reviews [2]. These 3307 negative reviews are further categorized into two classes, obscene and non-obscene. After annotation, this corpus consists of 2643 non-obscene reviews and 664 obscene reviews [3].

4 Creation of Obscene Lexicon

4.1 Text Processing Tools

4.1.1 POS Tagger

Part-Of-Speech (POS) tagger is a text analysis tool that assigns a POS tag (e.g., noun, verb, adjective, etc.) to each word of a piece of text data. As adjectives, nouns, and verbs usually convey opinions, the POS tagger can help to identify words that convey obscenity. Some of the popular POS taggers in English are NLTK POS tagger (Loper and Bird, 2002), spaCy POS tagger (Honnibal and Montani, 2017), etc.

We utilize a Bengali POS tagger to identify opinion conveying word (i.e., adjective and verb) [4]. However, the existing Bengali POS taggers are not as accurate as of its English counterpart. Hence, manual validation is needed to check the correctness of the POS tags assigned to words.

4.1.2 Word Embedding

The word embedding is a learned representation for textual content. A word embedding creates similar representations of words that are related in some ways. The word-embedding provides an efficient way to use the dense representation of words of varying lengths.

[2] https://github.com/sazzadcsedu/BN-Dataset

[3] https://github.com/sazzadcsedu/Bangla-vulgar-corpus.git

[4] https://github.com/AbuKaisar24/Bengali-Pos-Tagger-Using-Indian-Corpus/

Bengali	English Translation
তোমার মতো প্রতারক মাগিবাজ কম ই আছে।	There are fewer cheater whorehound like you.
তুই তো একটা মাদার চৌদ জামাত শিবিরের বেজন্মা	You are a motherfucker bastard Jamaat Shibir
গরম তেল পুটকির মধ্যে একটু ঢালেন কিছুটা হলে ও বুঝতে পারবেন	Pour a little hot oil into your asshole and you will understand
চুদির পো দিলি তো গান টা রে রেপ করে	Whore's son, you raped the song

Figure 1: Examples of obscene comments

There exist two main approaches for learning word embedding, count-based and context-based. The count-based vector space models heavily rely on the word frequency and co-occurrence matrix with the assumption that words in the same contexts share similar or related semantic meanings. The other learning approach, context-based methods, build predictive models that predict the target word given its neighbors. The best vector representation of each word is learned during the model training process.

The continuous Bag-of-Words (CBOW) model is a popular context-based method for learning word vectors. It predicts the center word from surrounding context words.

4.2 Lexicon Creation Framework

Leveraging lexical resources can assist in identifying the presence of profanity in Bengali social media. This study presents a semi-automatic approach for creating a swear lexicon utilizing an annotated corpus, word-embedding, and POS tagger. The lexicon development framework consisting of three phases, as shown in Figure 2,

1. Selection of seed words
2. Expansion of lexicon
3. Manual validation

4.2.1 Selection of Seed Words

The proposed methodology adopts a labeled obscene corpus to generate a list of seed words. The occurrences of individual words in the corpus are counted. Based on the word-occurrence count, the top 100 words are selected. We observe the presence of some non-vulgar words among the top 100 words, which are excluded.

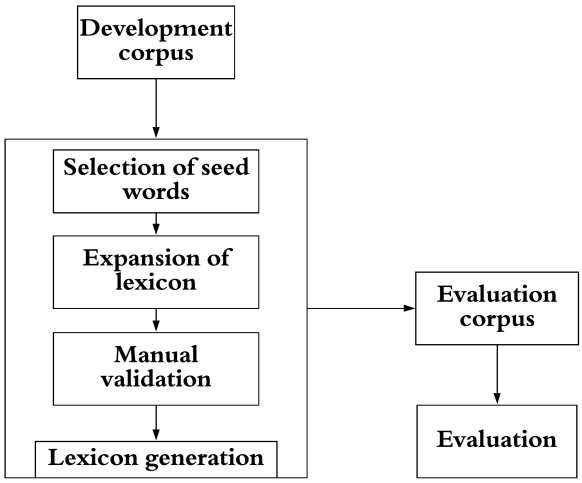

Figure 2: The lexicon creation framework

4.2.2 Expansion of Lexicon

The lexicon expansion step involves utilizing word embedding to identify similar words of the seed list. The Gensim (Řehůřek and Sojka, 2010) continuous Bag-of-Words (CBOW) implementation is used to find similar words in the *development corpus*.

The entire procedure consists of the following steps.

- In the first step, we identify words that are the most similar to the seed words.
- The second step iteratively finds words similar to obscene words recognized in the first step. The duplicate words are removed automatically. In addition, we exclude words that are not a noun, adjective, or verb. This iteration stops when we notice no significant expansion of the obscene word list.

4.2.3 Manual Validation

In the final step, we manually exclude non-obscene words that exist in the lexicon. As lexical resources

Bengali	English
খানকি	Whore
চোদা	Fuck
মাদারচোদ	Motherfucker
ভোদা	Pussy
বেশ্যা	Whore
গুদ	Pussy
গে	Gay
শুয়ার	Pigs
লেসবিয়ান	Lesbian
চুষে	Sucking
দুদ	Milk/Tits
পতিতা	Prostitute
জারজ	Bastard

Figure 3: Examples of obscene/swear words from the created lexicon

such as POS tagger in Bengali are not sophisticated enough, a manual validation step is necessary to eliminate non-obscene words. Moreover, we find that vulgar comments often do not follow the usual sentence structure; therefore, the POS tagger often fails to tag them correctly.

4.3 Developed Lexicon

Figure 3 shows some examples of Bengali obscene words and corresponding English translations. The English translation is provided to give an idea of the type/characteristics of the lexicon. Due to language-specific differences, the exact English translation may not be available for some of the words present in the developed obscene lexicon.

5 Obscenity Detection in Textual Content

As no lexicon exists in Bengali for obscene content detection, we compare the performance of the developed lexicon with several supervised classifiers in the evaluation corpus.

5.1 ML Classifiers

Two supervised ML classifiers, Logistic Regression (LR) and Support Vector Machine (SVM), and an optimizer, Stochastic Gradient Descendent (SGD), are employed in the evaluation corpus to identify profane reviews.

Logistic regression (LR) is a predictive analysis model that assigns observations into a discrete set of classes. LR assumes there are one or more independent variables that determine the outcome of the target.

Support Vector Machine (SVM) is a discriminative classifier defined by a separating hyperplane. Given the labeled training data, SVM generates an optimal hyperplane that categorizes unseen observations. For example, in two-dimensional space, this hyperplane is a line dividing a plane into two parts where each class lays on either side.

SGD is an optimization technique and does not correspond to a specific family of machine learning models. It is only a way to train a model. For SGD, we use hinge loss and l2 penalty.

As a feature vector, we use the unigram and bigram-based tf-idf score. 10-fold cross-validation is performed to assess the performance of various ML classifiers. We use the scikit-learn (Pedregosa et al., 2011) library implementation of the above-mentioned classifiers. For all the classifiers, default parameter settings are used.

To evaluate the performance of various ML classifiers and the SGD optimizer, we employ them in both original class-imbalanced and modified class-balanced settings. For class-balancing, initially, we employ the sampling method, SMOTE (Synthetic Minority Oversampling Technique) (Chawla et al., 2002). SMOTE is an oversampling technique where the synthetic samples are generated for the minority class with the help of interpolation. We notice it can not eliminate the inequality of class distribution in the original dataset. Thus, we use the subsampling method to make the dataset class-balanced.

In the original class-imbalanced setting, all the 664 obscene comments and 2643 non-obscene negative comments are utilized. In the modified class-balanced setting, all the 664 obscene comments are used; however, for the non-obscene class, we randomly select 664 non-obscene comments from a set of 2643 non-obscene comments.

5.2 Evaluation Metric

To show the effectiveness of the created lexicon, we utilize document-level coverage, $DCov$. $DCov$ of a lexicon corresponding to a review corpus is calculated as follows-

First, the number of reviews containing at least one word from the created lexicon is counted; afterward, it is divided by the total number of reviews present in the corpus. The following equation is used to calculate $DCov$ of a lexicon corresponding to a corpus-

Type	Method	Correctly Identified (Out of 664)	DCov/Recall
Lexicon-based	Developed Lexicon	564	0.8493
ML Classifier (Unbalanced)	LR	400	0.6024
	SVM	399	0.6009
	SGD	384	0.5783
ML Classifier (Balanced)	LR	609	0.9171
	SVM	594	0.8945
	SGD	589	0.8870

Table 1: Comparative Performance of Various Methods for Obscene Text Identification

$$DCov = \frac{\#reviews\ with\ (>0)\ obscene\ word\ identified}{total\ number\ of\ reviews\ in\ corpus}$$

The main motivation behind creating the obscene lexicon is to identify comments and reviews that contain swearing, profane, or obscene words; thus, $DCov$ is shown only for recognizing the usage of dirty language. Besides, the developed lexicon is manually validated at the final step; thus, it primarily contains obscene or profane words; hence, there is a very low chance that it identifies a non-obscene comment as obscene (false positive). However, this scenario could occur for few words present in the lexicon when they are used in different contexts. For example, one of the Bengali words in the lexicon may refer to either *tits* or *milk*, depending on the context.

For the ML classifiers, $DCov$ depicts the recall score for obscene class detection.

5.3 Comparison Results

Table 1 shows that among the 664 obscene reviews present in the evaluation corpus, the developed lexicon registers 564 reviews as obscene by recognizing the presence of at least one obscene/swear/profane term in the review, which is a document-level coverage of around 0.85.

Table 1 provides the coverage of various ML classifiers in the *evaluation corpus*. We present their performances in two different settings: original class-imbalanced setting and modified class-balanced setting.

From the Table 1, we observe that when the original class-imbalanced data is used, all the three ML classifiers achieve coverage of only 0.60. However, when a class-balanced dataset is utilized, the coverages of classifiers dramatically increase; they achieve around 0.90 coverage.

6 Discussion

The results reveal that the developed lexicon is capable of identifying obscene content in Bengali social media. It shows higher document-level coverage than in-domain labeled data in original class-imbalanced settings (i.e., when the dataset contains mostly non-obscene comments). However, in a class-balanced dataset, we find ML classifiers and the SGD optimizer perform better than the developed lexicon.

Labeled data are scarcely available in low-resource languages such as Bengali; therefore, the developed lexicon can be a practical resource for obscenity identification when labeled data are unavailable. Besides, as shown in Table 1, the performances of ML classifiers can be affected by the class distribution of the training dataset. The obscene or vulgar comments usually occupy only a small portion of a dataset. The ML classifiers can be less effective with the presence of a small sample size of obscene comments. The developed lexicon can be very effective in this scenario.

7 Summary and Conclusion

This study presents a semi-automatic methodology for creating a lexical resource (i.e., an obscene lexicon) to detect obscene content in Bengali. An obscene corpus and various text processing tools and resources are leveraged to develop the obscene lexicon. The developed lexicon is made publicly available [5]. The proposed methodologies can be adapted to other resource-limited languages to create lexical resources.

The efficacy of the obscene lexicon suggests that it can be utilized to distinguish obscenity in Bengali social media content when annotated data are

[5] https://github.com/sazzadcsedu/Bangla-Vulgar-Lexicon.git

unavailable. However, it should be noted that labeling textual content as obscene or vulgar entirely based on obscene or swear words may not be sufficient due to the complexity of the natural languages. Still, a well-annotated lexicon of moderate size can assist in identifying profane content; especially, in resource-scarce language. Our future work will involve expanding the size of the lexicon by utilizing larger and multi-domain development corpora.

Acknowledgments

The conference registration fee was supported by the ISAB VISA Scholarship of Old Dominion University.

References

2020. Abusive comment dataset. https://github.com/aimansnigdha/Bangla-Abusive-Comment-Dataset.git. Accessed 10 Sep. 2020.

Nayan Banik and Md Hasan Hafizur Rahman. 2019. Toxicity detection on bengali social media comments using supervised models. In *International Conference on Innovation in Engineering and Technology (ICIET)*, volume 23, page 24.

Tommaso Caselli, Valerio Basile, Jelena Mitrović, Inga Kartoziya, and Michael Granitzer. 2020. I feel offended, don't be abusive! implicit/explicit messages in offensive and abusive language. In *Proceedings of the 12th language resources and evaluation conference*, pages 6193–6202.

Puja Chakraborty and Md Hanif Seddiqui. 2019. Threat and abusive language detection on social media in bengali language. In *2019 1st International Conference on Advances in Science, Engineering and Robotics Technology (ICASERT)*, pages 1–6. IEEE.

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.

Michele Corazza, Stefano Menini, Elena Cabrio, Sara Tonelli, and Serena Villata. 2020. A multilingual evaluation for online hate speech detection. *ACM Transactions on Internet Technology (TOIT)*, 20(2):1–22.

Maral Dadvar, Dolf Trieschnigg, Roeland Ordelman, and Franciska de Jong. 2013. Improving cyberbullying detection with user context. In *European Conference on Information Retrieval*, pages 693–696. Springer.

Elisabeth Eder, Ulrike Krieg-Holz, and Udo Hahn. 2019. At the lower end of language—exploring the vulgar and obscene side of german. In *Proceedings of the Third Workshop on Abusive Language Online*, pages 119–128.

Estiak Ahmed Emon, Shihab Rahman, Joti Banarjee, Amit Kumar Das, and Tanni Mittra. 2019. A deep learning approach to detect abusive bengali text. In *2019 7th International Conference on Smart Computing & Communications (ICSCC)*, pages 1–5. IEEE.

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

Michael Gauthier, Adrien Guille, Fabien Rico, and Anthony Deseille. 2015. Text mining and twitter to analyze british swearing habits. *Handbook of Twitter for Research*.

Njagi Dennis Gitari, Zhang Zuping, Hanyurwimfura Damien, and Jun Long. 2015. A lexicon-based approach for hate speech detection. *International Journal of Multimedia and Ubiquitous Engineering*, 10(4):215–230.

Matthew Honnibal and Ines Montani. 2017. spacy 2: Natural language understanding with bloom embeddings, convolutional neural networks and incremental parsing. *To appear*, 7(1).

Alvi Md Ishmam and Sadia Sharmin. 2019. Hateful speech detection in public facebook pages for the bengali language. In *2019 18th IEEE International Conference On Machine Learning And Applications (ICMLA)*, pages 555–560. IEEE.

T. Jay and Kristin Janschewitz. 2008. The pragmatics of swearing.

Timothy Jay. 1992. *Cursing in America: A Psycholinguistic Study of Dirty Language in the Courts, in the Movies, in the Schoolyards, and on the Streets*. John Benjamins Publishing.

Md Karim, Bharathi Raja Chakravarthi, Mihael Arcan, John P McCrae, Michael Cochez, et al. 2020. Classification benchmarks for under-resourced bengali language based on multichannel convolutional-lstm network. *arXiv preprint arXiv:2004.07807*.

Edward Loper and Steven Bird. 2002. Nltk: the natural language toolkit. *arXiv preprint cs/0205028*.

Tony McEnery. 2004. *Swearing in English: Bad language, purity and power from 1586 to the present*. Routledge.

Chikashi Nobata, Joel Tetreault, Achint Thomas, Yashar Mehdad, and Yi Chang. 2016. Abusive language detection in online user content. In *Proceedings of the 25th international conference on world wide web*, pages 145–153.

Endang Wahyu Pamungkas, Valerio Basile, and Viviana Patti. 2020. Do you really want to hurt me? predicting abusive swearing in social media. In *The 12th Language Resources and Evaluation Conference*, pages 6237–6246. European Language Resources Association.

F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. 2011. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830.

Steven Pinker. 2007. *The stuff of thought: Language as a window into human nature*. Penguin.

Fabio Poletto, Valerio Basile, Manuela Sanguinetti, Cristina Bosco, and Viviana Patti. 2021. Resources and benchmark corpora for hate speech detection: a systematic review. *Language Resources and Evaluation*, 55(2):477–523.

Amir H Razavi, Diana Inkpen, Sasha Uritsky, and Stan Matwin. 2010. Offensive language detection using multi-level classification. In *Canadian Conference on Artificial Intelligence*, pages 16–27. Springer.

Radim Řehůřek and Petr Sojka. 2010. Software Framework for Topic Modelling with Large Corpora. In *Proceedings of the LREC 2010 Workshop on New Challenges for NLP Frameworks*, pages 45–50, Valletta, Malta. ELRA.

Salim Sazzed. 2020a. 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.

Salim Sazzed. 2020b. Development of sentiment lexicon in bengali utilizing corpus and cross-lingual resources. In *2020 IEEE 21st International Conference on Information Reuse and Integration for Data Science (IRI)*, pages 237–244. IEEE Computer Society.

Salim Sazzed. 2021a. Abusive content detection in transliterated bengali-english social media corpus. In *Proceedings of the Fifth Workshop on Computational Approaches to Linguistic Code-Switching*, pages 125–130.

Salim Sazzed. 2021b. Improving sentiment classification in low-resource bengali language utilizing cross-lingual self-supervised learning. In *International Conference on Applications of Natural Language to Information Systems*, pages 218–230. Springer.

Salim Sazzed and Sampath Jayarathna. 2019. A sentiment classification in bengali and machine translated english corpus. In *2019 IEEE 20th International Conference on Information Reuse and Integration for Data Science (IRI)*, pages 107–114. IEEE.

Salim Sazzed and Sampath Jayarathna. 2021. Ssentia: a self-supervised sentiment analyzer for classification from unlabeled data. *Machine Learning with Applications*, 4:100026.

Gary W Selnow. 1985. Sex differences in uses and perceptions of profanity. *Sex Roles*, 12(3-4):303–312.

Svitlana Volkova, Theresa Wilson, and David Yarowsky. 2013. Exploring demographic language variations to improve multilingual sentiment analysis in social media. In *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing*, pages 1815–1827, Seattle, Washington, USA. Association for Computational Linguistics.

Wenbo Wang, Lu Chen, Krishnaprasad Thirunarayan, and Amit P Sheth. 2014. Cursing in english on twitter. In *Proceedings of the 17th ACM conference on Computer supported cooperative work & social computing*, pages 415–425.

George RS Weir and Ana-Maria Duta. 2012. Strategies for neutralising sexually explicit language. In *2012 Third Cybercrime and Trustworthy Computing Workshop*, pages 66–74. IEEE.

Michael Wiegand, Josef Ruppenhofer, Anna Schmidt, and Clayton Greenberg. 2018. Inducing a lexicon of abusive words–a feature-based approach.

Ali Yadollahi, Ameneh Gholipour Shahraki, and Osmar R Zaiane. 2017. Current state of text sentiment analysis from opinion to emotion mining. *ACM Computing Surveys (CSUR)*, 50(2):1–33.

Yi Yang and Jacob Eisenstein. 2017. Overcoming language variation in sentiment analysis with social attention. *Transactions of the Association for Computational Linguistics*, 5:295–307.