Evaluating Machine Learning Algorithms For Bengali Fake News Detection

Conference Paper · December 2020

DOI: 10.1109/ICCIT51783.2020.9392662

4 authors, including:

Shafayat Bin Shabbir Mugdha
United International University
2 PUBLICATIONS **0** CITATIONS
SEE PROFILE

Sayeda Muntaha Ferdous
United International University
2 PUBLICATIONS **0** CITATIONS
SEE PROFILE

2020 23rd International Conference of Computer and
Information Technology (ICCIT), 19-21 December 2020

Evaluating Machine Learning Algorithms For Bengali Fake News Detection

Shafayat Bin Shabbir Mugdha, Sayeda Muntaha Ferdous, Ahmed Fahmin

Dept. of Computer Science and Engineering
United International University
Dhaka, Bangladesh

***Abstract*—In this world of modern technologies and media, online news publications and portals are increasing at a high speed. That is why, nowadays, it has become almost impossible to check out the traditional fact of news headlines and examine them due to the increase in the number of content writers, online media portals, and news portals. Mostly, fake headlines are filled with bogus or misleading content. They attract the commoners by putting phony words or misleading fraudulent content in the headlines to increase their views and share. But, these fake and misleading headlines create havoc in the commoner's life and misguide them in many ways. That is why we took a step so that the commoners can differentiate between fake and real news. We proposed a model that can successfully detect whether the story is fake or accurate based on the news headlines. We created a novel data set of Bengali language and achieved our aim and reached the target using the Gaussian Naive Bayes algorithm. We have used other algorithms, but the Gaussian Naive Algorithm has performed well in our model. This algorithm used a text feature dependent on TF-IDF and an Extra Tree Classifier to choose the attribute. In our model, using Gaussian Naive Bayes we got 87% accuracy which is comparatively best than any other algorithm we used in this model.**

***Index Terms*—Headlines, Machine Learning (ML), Natural Language Processing (NLP).**

I. Introduction

Headlines are the head or prominent or focused part at the top of the news, article, and content. This 'headline' part contains a summary of any content. That is why headlines are the most critical part of news or article writing. As headlines are the first thing that remains at the top of the news, so this plays a vital role in drawing attention and make the reader interested in the story quickly and briefly. Usually, headlines are written by the content or article writer.

As headlines are the first thing to grab the reader's attention, they play a vital role in our daily life. Sometimes, headlines with misleading, bogus, and full of fake content are published to sell the content and create many viewers. These misleading or fake headlines have become a burning question in our day to day life. As the number of media portals and content writers increases day by day, writers try to attract more people towards their news, article, or content. That's why, in the name of making the headlines attractive and eye-catching, they use inappropriate words, misleading and bogus contents. By doing so, sometimes, the commoners who don't read the full article falls into a trap, and they can't figure out the actual content or news. They start to spread that news from person to person, and this way, fake headlines get a high number of viewers. Not only that, by doing so, but commoners also face so many problems in their day to day life and pass through a tense situation or mental condition.

Researchers worldwide are trying to make a model that can detect fake news using full content. But several works have been attempting to catch the counterfeit using only the headlines. These kinds of attempts are minimal in number in the case of the Bengali Language. This paper tried to detect whether the news is real or fake based on news headlines. Usually, we can take two approaches to expose counterfeit headlines. They are Natural Language Processing and Fact-Checking Methods. To detect the Bengali Language's fake headlines, at first, we created a novel dataset in the Bengali Language. Among the above mentioned two processes, we chose Natural Language Processing (NLP) to build a method for our purpose. To the best of our knowledge, This work is the *first* to detect false headlines in the Bengali Language.

The fundamental *objective* of this work is to build a false headline detection system using machine learning. We also provide a comparison between our predicted results with actual outcomes. Our work covers: (i) comparing different feature selection algorithms, (ii) using an Extra Tree Classifier to select essential features, and (iii) testing on numerous algorithms in our dataset to find out the best one.

The remaining part of this paper is outlined as follows: In the Section II we discussed previous methods related to detect fake news. We propose more effective and practical approach based on NLP in Section III. Section V-A reports the results of our proposed model, followed by conclusions in Section VI.

978-0-7381-2333-2/20/$31.00 ©2020 IEEE

II. Related Works

Assessment against fake information disclosure is still at a fundamental stage, as this is a moderately progressing wonder, at any rate concerning the interest raised by society. We can sort fake news into three sorts. The first one is legitimate false news, which is altogether made up of nothing. The ensuing one is phony farce news, which is fake news whose driving job is to give humor. The third mix is insufficiently formed reports, which have some genuine news degree, yet they are not altogether exact.

Countless numbers of quantities of works have been executed to identify fake news or spam news over the recent few years. Social spam or phony story has pulled out the business's consideration because of the quick advancement of sites and online media stages.

A. Research over the last couple of years

The recognition of social spam [1] is probably the soonest paper ever. In that paper, six highlights of correlational labeling frameworks tuning different social spam classifiers properties were recognized by the creators, which accomplished an exactness above 98% by keeping up a false positive rate, which is 2%. A portion of the other late works [2]–[4] were likewise distributed concerning phony news discovery. The author [5] additionally proposed a mixed-race model. In [4], the creators discussed three distinct kinds of fake news. They are Exposed to manufactures, Large-scale deceptions, and finally, news parody. In [6] and [7], the creators had a conversation about the difficulties in the hour of working with fake news. In recent work, fake news detection was inspected by Devyani Keskar et al. dissected [8].

The creators clarified that the most recent development in natural language processing (NLP) and trickiness discovery could recognize fake news. In any case, the absence of accessible corpora for prescient demonstrating is a fundamental restricting component in making compelling models to recognize fake news.

Horne et al. [9] just spoke to the fact that it is so clear to separate among fake and real news. As per their perceptions, fake news titles have less stop-words and things. They separated various kinds of highlights arranged into three classes, as follows:

- Difficulty highlights determine the unpredictability and comprehensibility of the content.
- Frame of mind highlights represent and measure the intellectual cycle and individual concerns basic the works, for example, the number of feeling words and straightforward words.
- Elaborate features reflect the researchers' style and accentuation of the substance, for example, the number of action words and the number of things.

B. Different methods used over the last couple of research

The highlights, as referred to earlier, were utilized to fabricate an SVM classification model. Right when they considered real news against spoof articles (entertaining articles), they achieved 91% exactness. In any case, the accuracy dropped to 71% when anticipating fake news against real news.

Rubin et al. [10] recommended a model to recognize the satirical news analysis. They explored and assessed 360 satirical news stories in fundamentally four spaces: civics, science, business, and delicate news. They proposed an SVM classification model on their satirical news analysis. The five features are seen as Absurdity, Humor, Grammar, Negative effect, and Punctuation. The most critical exactness of 90% was accomplished, using only three blends of features, which are Absurdity, Grammar, and Punctuation.

There have been numerous sorts of classifiers used throughout the long term. RNN and LSTM were used in [11]. Wang et al. [12] used Logistic Regression, Support Vector Machine (SVM), Convolutional Neural Network (CNN), and Long short-term memory (LSTM) to detect fake news. Support Vector Machine (SVM) Classifier has been used in [10]. A basic methodology was proposed in the paper [7] using the Naive Bayes (NB) Classifier achieving 74% precision on the test set. A hybrid approach combining logistic regression and harmonious boolean label crowdsourcing was proposed in [13]. Geometrical deep learning was used in [14].

C. Different Datasets

In the literature, we can find diverse datasets. Zheng et al. [15] used a Chinese dataset collected from Weibo (a Chinese web-based media). This dataset classifies the user, two unique groups e.g., spammers and non-spammers. Tacchini et al. [13] used Facebook to gather the source of information. They used About 15,500 posts and user responses of around 909236 users to make their dataset. LIAR dataset has been used in many research works [11], [12]. There are two more popular datasets, one is the Twitter dataset [14], and the other one is the Buzzfeed dataset [16]. These datasets are generally used and acknowledged by researchers.

III. Modelling Approach

Fig.I shows our proposed methodology. Initially, in our procedures, we applied a novel strategy to collect a dataset. From that point moving, a preprocessing of the dataset has been finished by us to get stemmed sentences. We applied TF-IDF for feature extraction, and for feature selection, we applied Extra Tree Classifier. Lastly, we put them through classifiers for the accuracy rate. Various kinds of classifiers have been used in our dataset to detect fake news depending on the news headlines.

A. Dataset Creation

Bengali is the seventh most communicated in language on the planet, as per the authors of 'Ethnologue' [17]. Despite that, it is so difficult to find a definitive collection of the Bengali language. That is why we collected

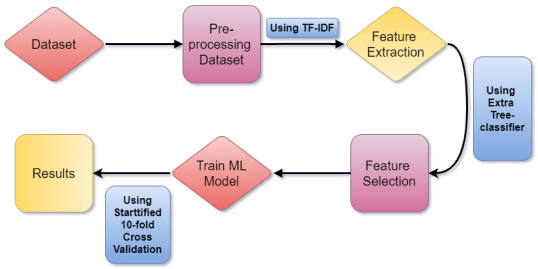

Fig. 1: Work Flow of Proposed Methodology

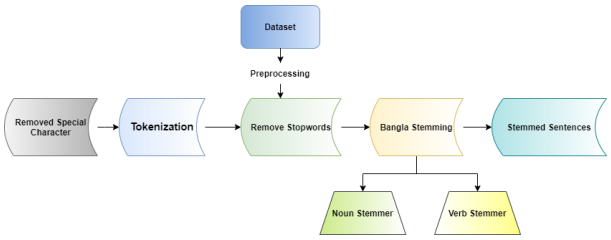

Fig. 2: Work Flow of Proposed Methodology

news to a great extent and used various connections and sources to develop our collection. We have made a point to incorporate information from multiple sites that contain fake news other than real stories, for example, BDFactCheck [18], Snopes.com, Fake News Detector (For Android Device), FactCheck.org, Politifact.com, etc. We have preferred news providers [19], [20] that distribute fake or misleading news now and again. We assembled the headlines, news URL, body, and the date. We labeled it fake or real, depending on their headline parts. We additionally looked through that news elsewhere in the world gateways, and on the off chance that it existed in those entries, after that, we expected it to be real news else we labeled it as fake news. Generally, we attempted to collect phony information relating to its actual report.

TABLE I: Dataset Visualization

News Headlines	Label
পরীক্ষা ছাড়াই অটোপ্রমোশনের খবর।	1
পঞ্চগড়ের তালমা নদীতে ভাঙন, গ্রামের রাস্তা নদীর গর্ভে বিলীন।	0
সাহাবউদ্দিন মেডিকেলের এমডি গ্রেফতার।	0
আইনমন্ত্রী করোনায় আক্রান্ত।	1
করোনাভাইরাস জানালা দিয়ে বের হতে পারে।	0

Table. I reveals a representation of our dataset. We have collected all the information from different reputable news entrances. The dataset has two columns, e.g., News Headlines and Label. There is a total of 538 instances. We separated the label into two categories. One is real, and another one is fake (Real: 269 and Fake: 269). For our research, we have only considered the news headlines.

B. Preprocessing

The preprocessing of our dataset is mainly done to extract a more refined dataset excluding stopwords and stemmers from the title feature of the dataset.

Our dataset's preprocessing steps are, for the most part, done to remove a more sifted dataset, barring stopwords, Tokenization, and stemmers from the title segment of the news, the preprocessing step functions as appeared in Fig 4.

1) Tokenization : In this progression, we disposed of all the special characters and the numeric values to have the texts to work with next. At that point, we divided the rest of the sentences into words, consequently tokenizing it from these sack words, we at that point eliminated the common stop words used every day [21]. Since filtering now, we have a lot of unique words. Fig 3 indicated the representation of Removing special characters and tokenizing sentences.

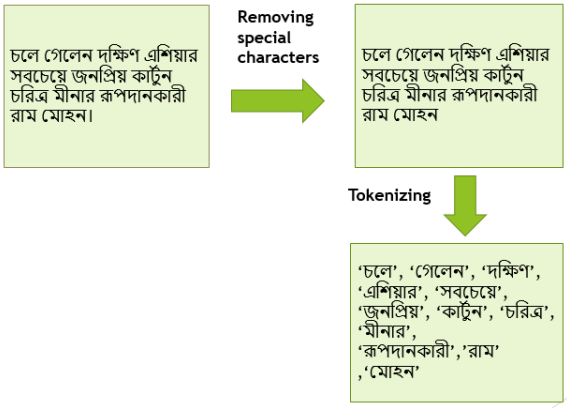

Fig. 3: visualization of Removing special characters and tokenizing sentences.

2) Stemming: : In the wake of tokenizing the data, the following stage changes over the tokens into a standard structure. Stemming changes the words in their original form and reduces the number of word types or classes. The collection of words that we got after tokenization used for stemming. The stemming has been done independently for the noun and the verb. The stemming is done similarly, as referenced in [22]. They did a 4 step verbal stemmer and 3 step noun stemmer, but our noun and verbal stemmer both consist of 3 steps:

- **Step 1:** We eliminated the inflected words.
- **Step 2:** We eliminated the diacritic mark from words.
- **Step 3:** Special cases dealt with, remembering a couple of changes for the diacritic mark for words.

INFLECTIONS IN BENGALI : Special cases are handled including a few transformation for the diacritic mark in the words.

- Verbal Inflections : An action word comprises of two sections, for example, verb = verb-root + verb sending. e.g., কর[kor] + এ[e] = করে [kore] Here, করে [kore]

TABLE II: Verbal Inflections.

Tense	1st & 2nd Person	2nd Person (Formal & Informal)	Formally (Honor)	Informally (Intimate)
Present Indefinite	ই[I]	এন[en]	এন[en]	এ[e]
Present Continuous	ছ[ch]	ছে[che], ছেন[chen]	ছেন[chen]	ছে [che]
Present Perfect	এছি [echi]	এছো [echo] এছেন [echen]	এছেন [echen]	এছে [eche]
Present Perfect Continuous	—	এন [en]	উন [un]	উক [uk]
Past Indefinite	লাম [lam]	লে [le], লেন [len]	লেন [len]	লা [la], ল [lo]
Past Continuous	ছিলাম [chilam]	ছিলে [chile], ছিলেন [chilen]	ছিলেন [chilen]	এছিলো [echilo]
Past Perfect	এছিলাম [echilam]	এছিলে [echile], এছিলেন [echilen]	এছিলেন [echilen]	এছিলো [echilo]
Habitual Past	তাম [tam]	তে [te], তেন [ten]	তেন [ten]	তা [ta], তো [to]
Habitual Future	বা [ba], ব [bo]	বে [be], বেন [ben]	বেন [ben]	বে [be]
Future Continuous	থাকবো [thakbo]	থাকবেন [thakben]	থাকবেন [thakben]	থাকবে [thakbe]
Future Perfect	থাকলো [thaklo]	থাকবে [thakbe]	থাকবেন [thakben]	থাকবে [thakbe]
Future Perfect Continuous	—	বেন [ben] , এন [en]	বেন [ben]	বে [be]

is verb, কর [kor] is the verb-root and এ [e] is the verb-ending.

- Noun Inflections : In Bengali, noun inflections happen because of various cases like nominative, evenhanded, genitive, and locative. These cases likewise contrast for singular and plural. Generally, singular thing expressions are shaped by the things finishing with রা [ra], টা [ta] টি [ti], খানা [khana], and so on and plural thing affectations are framed by the things finishing with এরা [era], গুলি [guli], গুলো [gulo] and so on. [23].

We used stemming for making classification faster and efficient. After the stemming, the remaining words merged to form a sentence known as Stemmed Sentences, which used for feature extraction.

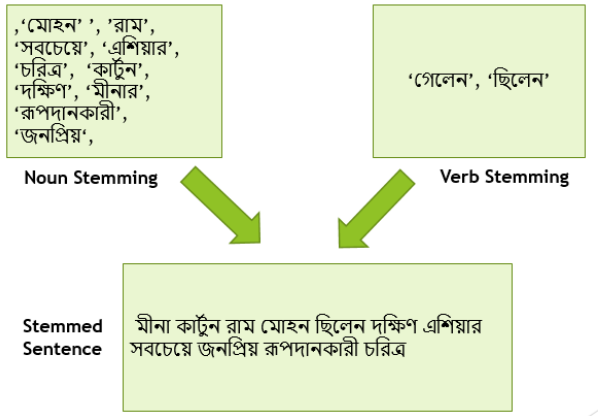

Fig. 4: Visualization of our Bangla Stemmer

IV. Feature Extraction & Selection

One of the difficulties of text classifiers is getting from high dimensional information. There are a few terms, words, and explanations in documents that lead to a high computational weight for the learning cycle. Besides, superfluous and excess highlights can harm the exactness and execution of the classifiers.

A. TF-IDF

Frequency–Inverserse Document Frequency known as TF-IDF is a commonly used method that uses transformed text numerical representation to determine how important a particular word in a document is. This is a widely used feature extraction technique for Natural Language Processing (NLP). One of IDF's principle qualities is that it impacts the term recurrence while moving up the uncommon ones. For instance, words, for example, "the" and "at that point" regularly show up in the content, and on the off chance that we just use TF, terms, for example, these will control the recurrence check. Nonetheless, utilizing the IDF measures down the effect of these terms.

B. Extra Tree Classifier

Extra Trees Classifier is a type of ensemble learning technique that aggregates various de-correlated decision trees collected in a "forest" to output its classification result. It is very similar to a Random Forest Classifier and only differs from it, informing the decision trees in the forest. In this case, we did not use it as a classifier, alternatively used it as a feature selection technique to select the best suitable features and then use the result in the Classifiers to get better results and better performance.

V. Experimental Analysis

All the experiments introduced that we have accomplished in this work on Jupyter notebook using Python 3.7 and Scikit-learn library. We have used stratified 10-fold cross-validation on the dataset to evaluate performance and nine algorithms that we have used in this work. Here, a few several performance metrics have been used. We have used accuracy, the area under the receiver Operating Characteristics curve (auROC), the area under the precision-Recall curve (auPR), F1 Score, and Matthews Correlation Coefficient (MCC).

A. Results

We have used multiple classification algorithms they are - Support Vector Machine (SVM), Logistic Regression (LR), multilayer perceptron (MLP), Random Forest Classifier (RF), VotingEnsemble Classifier (VEC), Gaussian Naive Bayes (GNB), Multinomial Naive Bayes (MNB), AdaBoost (AB) and Gradient Boosting (GB). Table III shows the performance of the classifiers with feature selection.

TABLE III: Performance of classifiers with Extra Tree

Classifiers	Accuracy	F1-Score	MCC	auROC	auPR
SVM (linear)	57.32%	0.721	0.211	0.621	0.538
LR	78.62%	0.691	0.413	0.775	0.821
MLP	72.93%	0.765	0.448	0.758	0.864
RF	61.14%	0.673	0.257	0.742	0.864
VEC	76.29%	0.772	0.545	0.857	0.891
GNB	87.42%	0.821	0.634	0.759	0.819
MNB	71.53%	0.625	0.378	0.736	0.834
AB	64.93%	0.675	0.408	0.721	0.764
GB	62.43%	0.712	0.323	0.822	0.734

B. Analysis

Emphasize that the results we have obtained for feature selection by using Extra Tree Classifier based feature selection is much better. Gaussian Naive Bayes classifier exceeds all other methods in terms of accuracy. In Fig 5 we can see the ROC curve for Gaussian Naive Bayes classifier. Besides these, in Fig 6 & Fig 7 we can see the word frequency of real and fake words.

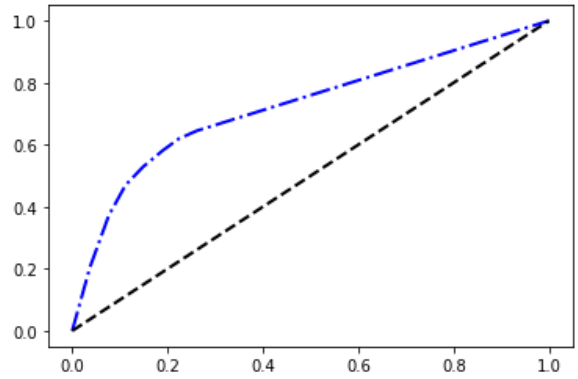

Fig. 5: ROC curve for Gaussian Naive Bayes

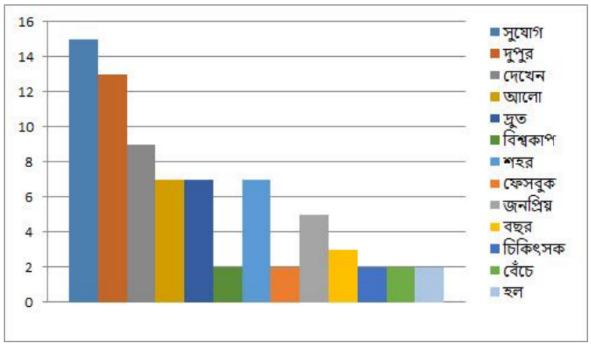

Fig. 6: Histogram showing Word Frequencies in Real Words

VI. Conclusion

In this research paper, we offered the Gaussian Naive Bayes algorithm for our model. This Gaussian Naive Bayes algorithm uses a text feature, dependent on TF-IDF and

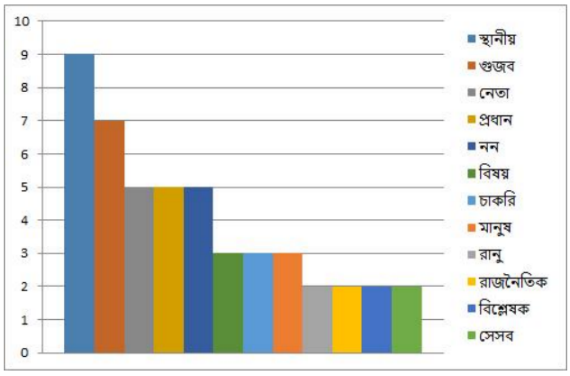

Fig. 7: Histogram showing Word Frequencies in Fake Words

an Extra Tree Classifier that chooses the feature. We created our very own novel dataset in the Bengali language and also a new stemmer. We reach our goal with the Gaussian Naive Bayes algorithm's help to detect whether the news is fake or real based on the news headlines. Hence, this study's results propose much more than artificial intelligence techniques that might be effectively used to handle this significant issue. In the future, we intend to develop a more robust and more effective model using more rich data so that it can extract features more precisely.

References

[1] B. Markines, C. Cattuto, and F. Menczer, "Social spam detection," in Proceedings of the 5th International Workshop on Adversarial Information Retrieval on the Web. ACM, 2009, pp. 41–48.

[2] V. L. Rubin, N. J. Conroy, and Y. Chen, "Towards news verification: Deception detection methods for news discourse," in Hawaii International Conference on System Sciences, 2015.

[3] N. J. Conroy, V. L. Rubin, and Y. Chen, "Automatic deception detection: Methods for finding fake news," Proceedings of the Association for Information Science and Technology, vol. 52, no. 1, pp. 1–4, 2015.

[4] V. L. Rubin, Y. Chen, and N. J. Conroy, "Deception detection for news: three types of fakes," in Proceedings of the 78th ASIS&T Annual Meeting: Information Science with Impact: Research in and for the Community. American Society for Information Science, 2015, p. 83.

[5] N. Ruchansky, S. Seo, and Y. Liu, "Csi: A hybrid deep model for fake news detection," in Proceedings of the 2017 ACM on Conference on Information and Knowledge Management. ACM, 2017, pp. 797–806.

[6] A. Figueira and L. Oliveira, "The current state of fake news: challenges ´ and opportunities," Procedia Computer Science, vol. 121, pp. 817–825, 2017.

[7] X. Zhou, R. Zafarani, K. Shu, and H. Liu, "Fake news: Fundamental theories, detection strategies and challenges," in Proceedings of the Twelfth ACM International Conference on Web Search and Data Mining. ACM, 2019, pp. 836–837.

[8] D. Keskar, S. Palwe, and A. Gupta, "Fake news classification on twitter using flume, n-gram analysis, and decision tree machine learning technique," in Proceeding of International Conference on Computational Science and Applications. Springer, 2020, pp. 139–147.

[9] Horne, B.D., Adali, S.: This just in: fake news packs a lot in the title, uses simpler, repetitive content in the text body, more similar to satire than real news. In: the 2nd International Workshop on News and Public Opinion at ICWSM (2017)

[10] Rubin., Victoria, L., et al.: Fake news or truth? Using satirical cues to detect potentially misleading news. In: Proceedings of NAACL-HLT (2016)
[11] S. Girgis, E. Amer, and M. Gadallah, "Deep learning algorithms for detecting fake news in online text," in the 2018 13th International Conference on Computer Engineering and Systems (ICCES). IEEE, 2018, pp. 93–97.
[12] W. Y. Wang, "liar, liar pants on fire": A new benchmark dataset for fake news detection," arXiv preprint ar X iv:1705.00648, 2017.
[13] E. Tacchini, G. Ballarin, M. L. Della Vedova, S. Moret, and L. de Alfaro, "Some like it hoax: Automated fake news detection in social networks," arXiv preprint ar X iv:1704.07506, 2017.
[14] F. Monti, F. Frasca, D. Eynard, D. Mannion, and M. M. Bronstein, "Fake news detection on social media using geometric deep learning," arXiv preprint arXiv:1902.06673, 2019.
[15] X. Zheng, Z. Zeng, Z. Chen, Y. Yu, and C. Rong, "Detecting spammers on social networks," Neurocomputing, vol. 159, pp. 27–34, 2015.
[16] M. Potthast, J. Kiesel, K. Reinartz, J. Bevendorff, and B. Stein, "A stylometric inquiry into hyperpartisan and fake news," arXiv preprint arXiv:1702.05638, 2017.
[17] Ethnologue, "List of 200 most spoken languages," 2020 (accessed January 15, 2020).
[18] "Bdfactcheck," 2020 (accessed January 15, 2020). [Online]. Available: https://www.bdfactcheck.com/
[19] "earki," 2020 (accessed January 15, 2020). [Online]. Available: https://www.earki.com/
[20] "Crazzfeed," 2020 (accessed January 15, 2020). [Online]. Available: https://crazzfeed.com/
[21] "Bengali stopwords," 2020 (accessed January 15, 2020). [Online]. Available: https://www.ranks.nl/stopwords/bengali.
[22] M. R. Mahmud, M. Afrin, M. A. Razzaque, E. Miller, and J. Iwashige, "A rule-based Bengali stemmer," in the 2014 International Conference on Advances in Computing, Communications, and Informatics (ICACCI). IEEE, 2014, pp. 2750–2756.
[23] H. Ruth Thompson, "Bengali - A comprehensive grammar."Published by routledge, 2 park square, Miltonpark, United Kingdom, 2010.