An Annotated Bangla Sentiment Analysis Corpus

Fuad Rahman
Apurba Technologies Ltd.
Dhaka, Bangladesh
fuad@apurbatech.com

Habibur Khan
Apurba Technologies Ltd.
Dhaka, Bangladesh
habib@apurbatech.com

Zakir Hossain
Apurba Technologies Ltd.
Dhaka, Bangladesh
zakir@apurbatech.com

Mahfuza Begum
Apurba Technologies Ltd.
Dhaka, Bangladesh
mahfuza@apurbatech.com

Sadia Mahanaz
Apurba Technologies Ltd.
Dhaka, Bangladesh
sadia@apurbatech.com

Ashraful Islam
Apurba Technologies Ltd.
Dhaka, Bangladesh
ashraful@apurbatech.com

Aminul Islam
Apurba Technologies Ltd.
Dhaka, Bangladesh
aminul@apurbatech.com

2019 International Conference on Bangla Speech and Language Processing (ICBSLP) 978-1-7281-5241-7/20/$31.00 ©2020 IEEE 10.1109/ICBSLP47725.2019.201474

978-1-7281-5241-7/19/$31.00 ©2019 IEEE

***Abstract* – This paper presents a Bangla corpus specifically targeted for sentiment analysis and made available to researchers under an open-source licensing scheme[1]. We have collected and manually annotated over 10,000 sentences with sentiment polarity. We then moved to the Word domain and annotated over 15,000 words derived from these sentences with sentiment polarity. Each entry is the corpus has been cross-annotated by at least two and sometimes three annotators for ensuring quality. Also as a pre-requisite of creating a high quality sentiment analysis corpus, we had to build a secondary corpus for Bangla word stemming, which is also been cross-validated by at least two and sometimes three annotators for ensuring quality.**

Index Terms – Sentiment Analysis, NLP, Bangla Corpus, Annotation, Open Source Corpus

I. Introduction

Sentiment analysis is a very important part of natural language processing. While very robust solutions for English already exists both in academic and commercial domains, for Bangla language, work in this area is still in its infancy. As the focus of tools for sentiment analysis has now shifted from rule based to machine learning methods, the need for annotated and ground truth data for training these solutions are of utmost importance. Unfortunately there is almost no serious corpus for Bangla language that is available for sentiment analysis, forcing researchers to stich together their own small corpora which are neither standardized and nor rigorously quality controlled. In this paper, we present a fully annotated corpus for sentiment analysis.

A. Brief Background

In recent times, there has been some research reported on Bangla sentiment analysis. One common approach seems to be to translate a Bangla word and then use the polarity from the English translated word. [1][6][7]. While it works on straightforward words, it cannot handle the nuances of a language. For example the word "জটিল" means "complex" [2] and it has a negative sense. But in Bangla it is often used in a positive sense, for example, "তামিম আজ জটিল খেলছে". Another example is the word "খাওয়া" which is commonly translated to "eat", the common polarity of which is neutral. But in the sentence "তার খাওয়া নাই", the polarity is distinctly negative.

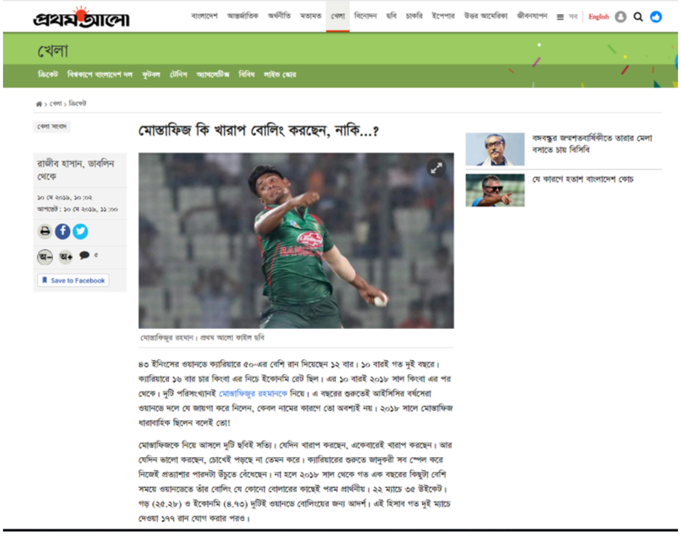
The reason for the popularity of this type of approach is very simple, a distinct lack of a ground truth corpus suited for training machine learning algorithms. Although there are some existing data set for sentiment analysis, but most of these are not available publicly. Some publicly available data set are small, e.g. [4] has about 4,000 sentences, whereas [1] has about 7,000 sentences.

One of the most significant resources is described in [6]. This corpus size is around 10,000 sentences. These sentences were collected from Facebook, Twitter, YouTube, online news portals and product review pages.

II. Sentiment Analysis Corpus

A. Data Source

The source of this data is the online Sports section of Prothom Alo, as shown in Fig 1 below.



মোস্তাফিজ কি খারাপ বোলিং করছেন, নাকি...?

রাজীব হাসান, ডাবলিন থেকে

১০ মে ২০১৯, ১০:০২
আপডেট: ১০ মে ২০১৯, ১১:০০

মোস্তাফিজুর রহমান। প্রথম আলো ফাইল ছবি

৪৩ ইনিংসের ওয়ানডে ক্যারিয়ারে ৫০-এর বেশি রান দিয়েছেন ১২ বার। ১০ বারই গত দুই বছরে। ক্যারিয়ারে ১৬ বার চার কিংবা এর নিচে ইকোনমি রেট ছিল। এর ১০ বারই ২০১৮ সাল কিংবা এর পর থেকে। দুটি পরিসংখ্যানই মোস্তাফিজুর রহমানকে নিয়ে। এ বছরের শুরুতেই আইসিসির বর্ষসেরা ওয়ানডে দলে যে জায়গা করে নিলেন, কেবল নামের কারণে তো অবশ্যই নয়। ২০১৮ সালে মোস্তাফিজ ধারাবাহিক ছিলেন বলেই তো!

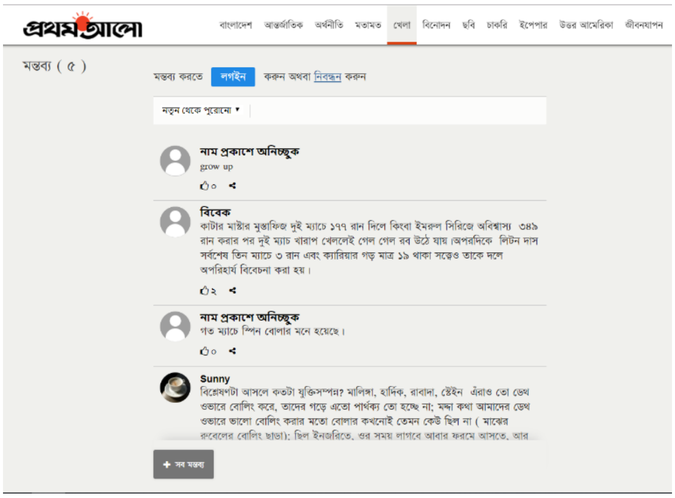
মোস্তাফিজকে নিয়ে আসলে দুটি ছবিই সত্যি। যেদিন খারাপ করছেন, একেবারেই খারাপ করছেন। আর যেদিন ভালো করছেন, চোখেই পড়ছে না তেমন করে। ক্যারিয়ারের শুরুতে জাদুকরী সব স্পেল করে নিজেই প্রত্যাশার পারদটা উঁচুতে বেঁধেছেন। না হলে ২০১৮ সাল থেকে গত এক বছরের কিছুটা বেশি সময়ে ওয়ানডেতে তাঁর বোলিং যে কোনো বোলারের কাছেই পরম প্রার্থনীয়। ২২ ম্যাচে ৩৫ উইকেট। গড় (২৫.২৮) ও ইকোনমি (৪.৭৩) দুটিই ওয়ানডে বোলিংয়ের জন্য আদর্শ। এই হিসাব গত দুই ম্যাচে দেওয়া ১৭৭ রান যোগ করার পরও।

বঙ্গবন্ধুর জন্মশতবার্ষিকীতে তারার মেলা বসাতে চায় বিসিবি

যে কারণে হতাশ বাংলাদেশ কোচ

Fig. 1 Daily Prothom Alo Online Edition.

Most of source sentences for the corpus were collected from the comments Section as shown in Fig 2 below.



মন্তব্য (৫)

মন্তব্য করতে লগইন করুন অথবা নিবন্ধন করুন

নাম প্রকাশে অনিচ্ছুক
grow up
০

বিবেক
কাটার মাষ্টার মুস্তাফিজ দুই ম্যাচে ১৭৭ রান দিলে কিংবা ইমরুল সিরিজে অবিশ্বাস্য ৩৪৯ রান করার পর দুই ম্যাচ খারাপ খেললেই গেল গেল রব উঠে যায় ।অপরদিকে লিটন দাস সর্বশেষ তিন ম্যাচে ৩ রান এবং ক্যারিয়ার গড় মাত্র ১৯ থাকা সত্ত্বেও তাকে দলে অপরিহার্য বিবেচনা করা হয় ।
২

নাম প্রকাশে অনিচ্ছুক
গত ম্যাচে স্পিন বোলার মনে হয়েছে।
০

Sunny
বিশ্লেষণটা আসলে কতটা যুক্তিসম্পন্ন? মালিঙ্গা, হার্দিক, রাবাদা, স্টেইন এঁরাও তো ডেথ ওভারে বোলিং করে, তাদের গড়ে এতো পার্থক্য তো হচ্ছে না; মন্দা কথা আমাদের ডেথ ওভারে ভালো বোলিং করার মতো বোলার কখনোই তেমন কেউ ছিল না (মাঝের রুবেলের বোলিং ছাড়া); ছিল ইনজুরিতে, ওর সময় লাগবে আবার ফরমে আসতে, আর

Fig. 2 Comments in the Sports Section.

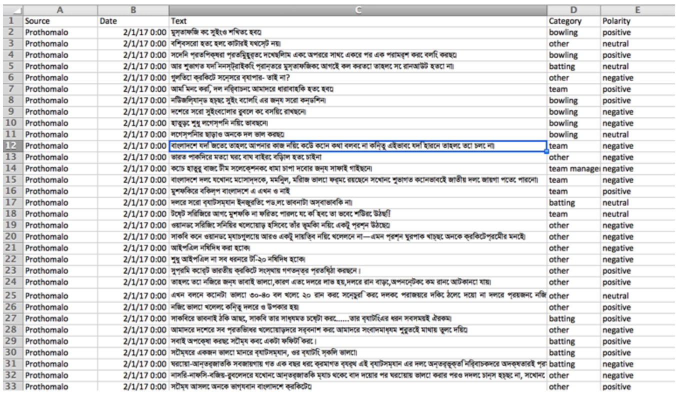
B. Methodology

The corpus was prepared in a combination of manual and automated steps. Initially the sentences appearing in the Sports Section were copied manually. These data were labeled by hand into three categories "positive", "negative" and "neutral", by a "Content Team" and then crosschecked by a "QA Team".

The truth labeling is done at two levels.

[1] See Section VI for details

Sentence Level: The first level is the sentence level, as seen in Fig 3 below. In this case, the polarity applies to the overall sentence.



	A	B	C	D	E
1	Source	Date	Text	Category	Polarity
2	Prothomalo	2/1/17 0:00	মুস্তাফিজ কে সুইংও শিখতে হবে।	bowling	positive
3	Prothomalo	2/1/17 0:00	বিশ্বসেরা হতে হলে কাটারই যথেস্ট নয়।	other	neutral
4	Prothomalo	2/1/17 0:00	সেদিন প্রতিপক্ষরা প্রতিমুহূর্তে দেখেছিলাম একে অপরের সাথে একের পর এক পরামর্শ করে বলিং করছে।	bowling	positive
5	Prothomalo	2/1/17 0:00	আর শুভাগত যদি নিনস্ট্রাইকিং প্রান্তের মুস্তাফিজকে আগেই কল করতো তাহলে সে রানআউট হতো না।	batting	neutral
6	Prothomalo	2/1/17 0:00	গুলতি ক্রিকেটে সন্দেহের ব্যাপার- তাই না?	other	negative
7	Prothomalo	2/1/17 0:00	আমি মনে করি, দল নির্বাচনে আমাদের ধারাবাহিক হতে হবে।	team	positive
8	Prothomalo	2/1/17 0:00	নিউজিল্যান্ড হচ্ছে সুইং বোলিং এর জন্য সেরা কন্ডিশন।	bowling	positive
9	Prothomalo	2/1/17 0:00	দেশের সেরা সুইংবোলার বুবলে কে বসিয়ে রাখছেন।	bowling	negative
10	Prothomalo	2/1/17 0:00	হাতুড়ে শুধু লেগস্পিন নিয়ে ভাবছেন।	bowling	negative
11	Prothomalo	2/1/17 0:00	লেগস্পিনার ছাড়াও অনেক দল ভাল করছে।	bowling	neutral
12	Prothomalo	2/1/17 0:00	বাংলাদেশ যদি জিতে তাহলে আপনার কাজ নিয়ে কেউ কোন কথা বলবে না কিন্তু এইভাবে যদি হারেন তাহলে তো চলে না।	team	negative
13	Prothomalo	2/1/17 0:00	ভারত পাকিদের মতো ঘরে বাঘ বাইরে বিড়াল হতে চাইনা	other	negative
14	Prothomalo	2/1/17 0:00	কোচ হাতুরু বাজে টীম সিলেক্শনকে ধামা চাপা দেবার জন্য সাফাই গাইছেন।	team manage	negative
15	Prothomalo	2/1/17 0:00	বাংলাদেশ দলে যেখানে মোসাদ্দেককে, মমিনুল, মিরাজ ভালো ফর্মে রয়েছেন সেখানে শুভাগত কোনভাবেই জাতীয় দলে জায়গা পেতে পারেনা।	team	negative
16	Prothomalo	2/1/17 0:00	মুশফিকের বিকল্প বাংলাদেশ এ এখন ও নাই	team	positive
17	Prothomalo	2/1/17 0:00	দলের সেরা ব্যাটসম্যান ইনজুরিতে পড়লে ভাবনাটা অস্বাভাবিক না।	batting	neutral
18	Prothomalo	2/1/17 0:00	টেস্ট সিরিজের আগে মুশফিক না ফিরতে পারলে যে কি হবে তা ভেবে শঙ্কিত উঠছি!	team	neutral
19	Prothomalo	2/1/17 0:00	ওয়ানডে সিরিজে সিনিয়র খেলোয়াড় হিসেবে তাঁর ভূমিকা নিয়ে একটু প্রশ্ন উঠছে।	other	negative
20	Prothomalo	2/1/17 0:00	সাকিব কেন ওয়ানডে ম্যাচগুলোয় আরও একটু দায়িত্ব নিয়ে খেললেন না—এমন প্রশ্ন ঘুরপাক খাচ্ছে অনেক ক্রিকেটপ্রেমীর মনেই।	other	negative
21	Prothomalo	2/1/17 0:00	আইপিএল নিষিদ্ধ করা হোক।	other	negative
22	Prothomalo	2/1/17 0:00	শুধু আইপিএল না সব ধরনের টি-২০ নিষিদ্ধ হোক।	other	negative
23	Prothomalo	2/1/17 0:00	সুপ্রিম কোর্ট ভারতীয় ক্রিকেট সংস্থায় গণতন্ত্র প্রতিষ্ঠা করছেন।	other	positive
24	Prothomalo	2/1/17 0:00	তাহলে তো নিজের জন্য ভাবাই ভালো,কারণ এতে দলের লাভ হয়,দলের রান বাড়ে,অপনেন্টকে কম রানে আটকানো যায়।	other	positive
25	Prothomalo	2/1/17 0:00	এখন বলেন কোনটা ভালো ৩০-৪০ বল খেলে ২০ রান করে সেন্চুরি করে দলকে পরাজয়ের দিকে ঠেলে দেয়া না দলের প্রয়োজনে নিজ	other	neutral
26	Prothomalo	2/1/17 0:00	নিজে ভালো খেললে কিন্তু দলের ও উপকার হয়।	other	positive
27	Prothomalo	2/1/17 0:00	সাকিবের ভাবনাই ঠিক আছে, সাকিব তার সাধ্যমত চেষ্টা করে......তার ব্যাটিংএর ধরন সবসময়ই এরকম।	batting	positive
28	Prothomalo	2/1/17 0:00	আমাদের দেশের সব প্রতিভাধর খেলোয়াড়দের সর্বনাশ করে আমাদের সংবাদমাধ্যম শুরুতেই মাথায় তুলে দিয়ে।	other	negative
29	Prothomalo	2/1/17 0:00	সবাই অপেক্ষা করছে সৌম্য কবে একটা ফিফটি করে।	batting	positive
30	Prothomalo	2/1/17 0:00	সৌম্যরে একজন ভালো মানের ব্যাটসম্যান, ওর ব্যাটিং স্কিল ভালো।	batting	positive
31	Prothomalo	2/1/17 0:00	ঘরোয়া-আন্তর্জাতিক সবজায়গায় গত এক বছর ধরে ক্রমাগত ব্যর্থ এই ব্যাটসম্যান এর দলে অন্তর্ভুক্তি নির্বাচকদের অদক্ষতারই প্র	batting	negative
32	Prothomalo	2/1/17 0:00	নাসির-নাফিস-বিজয়-বুবলেদের যেখানে আন্তর্জাতিক ম্যাচ থেকে বাদ দেয়ার পর ঘরোয়ায় ভালো করার পরও দলে চান্স হচ্ছে না, সেখানে	other	negative
33	Prothomalo	2/1/17 0:00	সৌম্য আসলে অনেক ভাগ্যবান বাংলাদেশ ক্রিকেটে।	other	positive

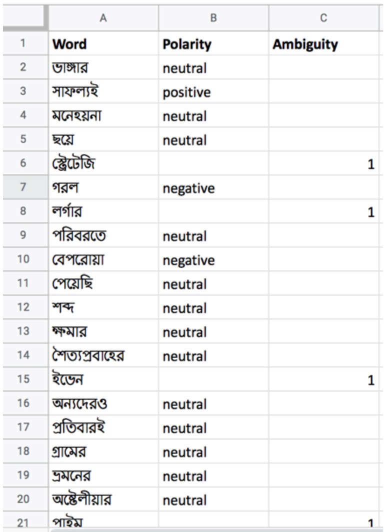
Fig. 3 Sentence level polarity.

- Manual collection and preprocessing: We collected more than 10,000 sentences from comments from online Bangla newspaper (https://www.prothomalo.com/), primarily from cricket sports news. We only included valid and complete single sentences. The task was distributed among the 5 members team members, who are all native Bangla speakers. Each member tagged sentiment polarity for the sentences allocated to him/her. Another member then crosschecked this. A third member then re-assigned the polarity sentences if the first and second members disagreed. The same methodology was applied to crosschecking sentence validation. In case there are disagreements, the assigned final polarity is at least assigned by two team members. If all three team members disagreed, the sentence was considered to be too ambiguous and dropped from the corpus.
- Automatic processing: We removed unwanted characters, words and symbols from the sentences, such as:
  - ['১','২','৩','৪','৫','৬','৭','৮','৯','1','2','3','4','5','6','7','8','9']
  - ['□','?',',',';',':','.','(','-','–','/','_','*','%','!','\'','+','<','>','—','০','=']
  - ['’','"','|','…',')','`','@','#','।','“','&','–','_','😚','🎐','👌','😆','😂']
  - [A-Z]
  - [a-z]

  We also removed duplicate sentences.

Word Level: The second is polarity on the word level, as shown in Fig 4 below.

- Automatic processing: We tokenized all collected sentences and removed numbers, digits, and symbols from the tokenized word list. We then identified the unique words from the word list of nearly 15,000 words. We stemmed the word list using two different stemmers, StemmerR[10] and StemmerP[11]. We then identified the words that produced the same result from these two stemmers.

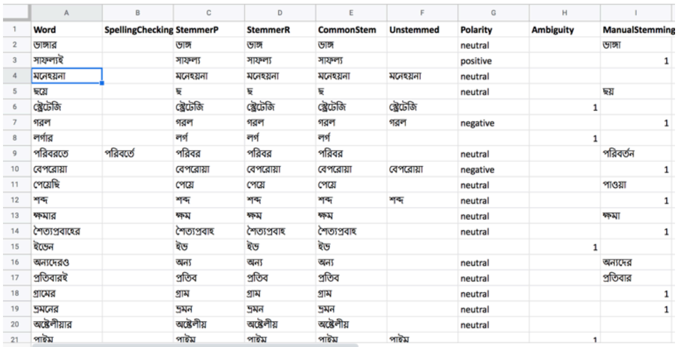


	A	B	C
1	Word	Polarity	Ambiguity
2	ভাঙ্গার	neutral	
3	সাফল্যই	positive	
4	মনেহয়না	neutral	
5	ছয়ে	neutral	
6	স্ট্রেটেজি		1
7	গরল	negative	
8	লর্গার		1
9	পরিবরতে	neutral	
10	বেপরোয়া	negative	
11	পেয়েছি	neutral	
12	শব্দ	neutral	
13	ক্ষমার	neutral	
14	শৈত্যপ্রবাহের	neutral	
15	ইডেন		1
16	অন্যদেরও	neutral	
17	প্রতিবারই	neutral	
18	গ্রামের	neutral	
19	ভ্রমনের	neutral	
20	অস্ট্রেলীয়ার	neutral	
21	পাইম		1

Fig. 4 Word level polarity.

- Manual collection and preprocessing: We checked whether the un-stemmed word is already a root or not and manually corrected the roots for those words that were stemmed wrongly. Once a clean word list was created, we then tagged the polarity of each word manually, using the same three-tiered approach as described before. This step also resulted in identifying some words that were ambiguous. These are then dropped for the final corpus.

A snapshot of this corpus is shows in Fig 5 below.



	A	B	C	D	E	F	G	H	I
1	Word	SpellingChecking	StemmerP	StemmerR	CommonStem	Unstemmed	Polarity	Ambiguity	ManualStemming
2	ভাঙ্গার		ভাঙ্গ	ভাঙ্গ	ভাঙ্গ		neutral		ভাঙ্গা
3	সাফল্যই		সাফল্য	সাফল্য	সাফল্য		positive		1
4	মনেহয়না		মনেহয়না	মনেহয়না	মনেহয়না	মনেহয়না	neutral		
5	ছয়ে		ছ	ছ	ছ		neutral		ছয়
6	স্ট্রেটেজি		স্ট্রেটেজি	স্ট্রেটেজি	স্ট্রেটেজি	স্ট্রেটেজি		1	
7	গরল		গরল	গরল	গরল	গরল	negative		1
8	লর্গার		লর্গ	লর্গ	লর্গ			1	
9	পরিবরতে	পরিবর্তে	পরিবর	পরিবর	পরিবর		neutral		পরিবর্তন
10	বেপরোয়া		বেপরোয়া	বেপরোয়া	বেপরোয়া	বেপরোয়া	negative		1
11	পেয়েছি		পেয়ে	পেয়ে	পেয়ে		neutral		পাওয়া
12	শব্দ		শব্দ	শব্দ	শব্দ	শব্দ	neutral		1
13	ক্ষমার		ক্ষম	ক্ষম	ক্ষম		neutral		ক্ষমা
14	শৈত্যপ্রবাহের		শৈত্যপ্রবাহ	শৈত্যপ্রবাহ	শৈত্যপ্রবাহ		neutral		1
15	ইডেন		ইড	ইড	ইড			1	
16	অন্যদেরও		অন্য	অন্য	অন্য		neutral		অন্যদের
17	প্রতিবারই		প্রতিব	প্রতিব	প্রতিব		neutral		প্রতিবার
18	গ্রামের		গ্রাম	গ্রাম	গ্রাম		neutral		1
19	ভ্রমনের		ভ্রমন	ভ্রমন	ভ্রমন		neutral		1
20	অস্ট্রেলীয়ার		অস্ট্রেলীয়	অস্ট্রেলীয়	অস্ট্রেলীয়		neutral		
21	পাইম		পাইম	পাইম	পাইম	পাইম		1	

Fig. 5 Corpus after stemming using two different algorithms.

III. Corpus Statistics

This Section qualifies the corpus.

TABLE I
RAW DATA COLLECTED FOR THE SENTIMENT CORPUS

Total number of sentences	10,008
Total number of words before filtering	19,731

Total number of words after filtering	14,874
How many ambiguous	2,140
How many words accepted	12,734

Table 1 shows the statistics of raw data collected from the sentiment corpus.

TABLE 2
COMPARING THE TWO STEMMERS

	StemmerP	StemmerR
How many words were stemmed	14,874	14,874
How many times the stemmer was able to stem a word	14,662	14,874
How many times the stemmer was not able to stem a word	212	14,874

Table 2 compares the performance of the two stemmers. This step was completely manual and three-level peer-viewed.

We found that 1.5% times the stemmers did not agree with each other. In order to correct that, we manually fixed the stemming.

TABLE 3
MANUAL STEMMING

Category	Number
Accuracy of StemmerP	58.08%
Accuracy of StemmerR	59.65%
How many words were manually stemmed	5,138
How many spelling were corrected manually	682

Table 3 shows the results of manually fixing stemming issues.

After all this cleanup and manual fixes, the final corpus has the following entries, as seen in Table 4.

TABLE I
RAW DATA COLLECTED FOR THE SENTIMENT CORPUS

	Positive		Negative		Neutral	
	Actual	%	Actual	%	Actual	%
Total Number of sentences	3,183	33.0	4,110	42.67	2,337	24.27
Total Number of words	824	6.47	1,068	8.38	10,804	84.80

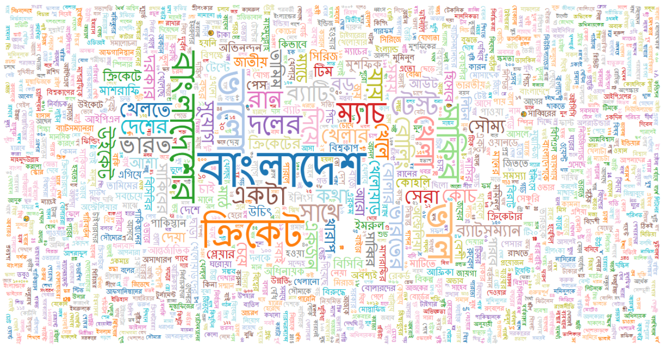


Fig. 6 WordCloud of corpus sentences with both positive and negative polarity.

IV. CORPUS PROPERTIES

A. *WordCloud of Collected Sentences*

Fig. 6 shows a WordCloud of the collected sentences with positive and negative polarity. Please note that we have used a stop word list to filter the words before this was created. We have identified 398 stop words. The same applies to building WordClouds for negative and neutral sentences.

Fig. 7 shows a WordCloud of the collected sentences with positive polarity.

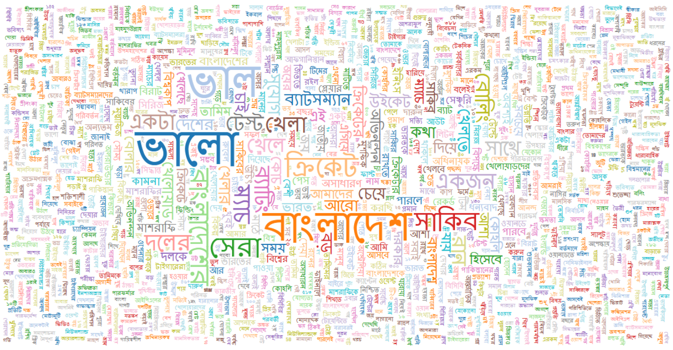


Fig. 7 WordCloud of corpus sentences with a positive polarity.

Fig. 8 shows a WordCloud of the collected sentences with negative polarity.

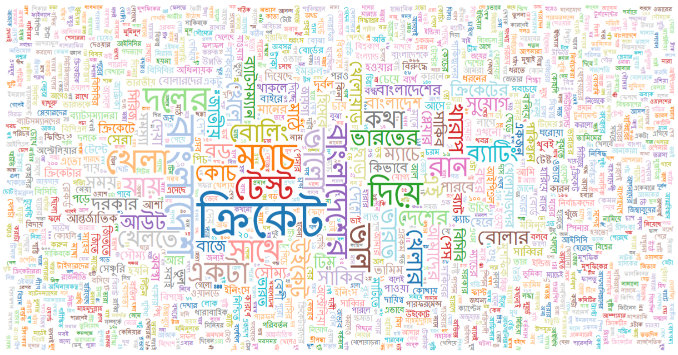


Fig. 8 WordCloud of corpus sentences with a negative polarity.

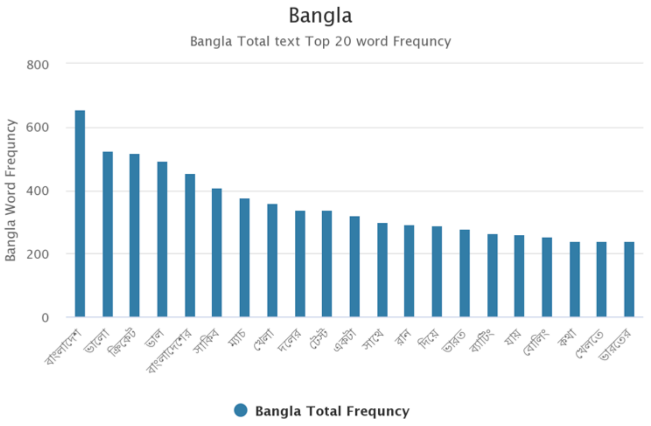


Fig. 9 Word frequency of top 20 words

B. Frequency Distribution of Top 20 Words

Fig. 9 shows the frequency of the top 20 words in our corpus.

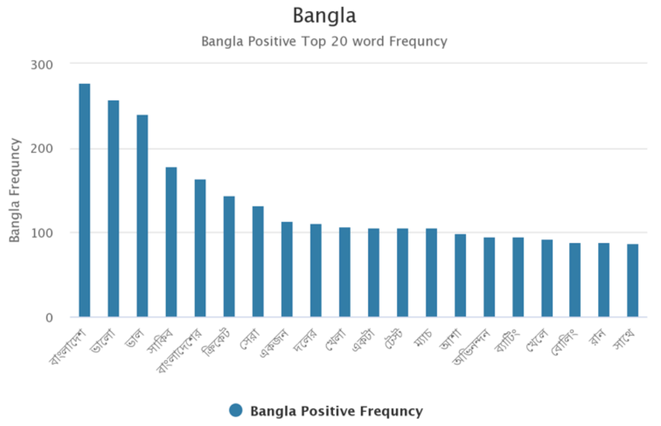


Fig. 10 Word frequency of top 20 positive words

Fig. 10 shows the frequency of the top 20 positive words in our corpus.

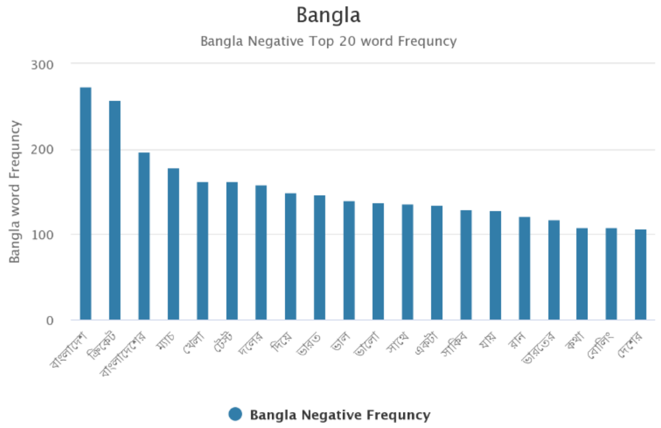


Fig. 11 Word frequency of top 20 negative words

Fig. 11 shows the frequency of the top 20 negative words in our corpus.

V. Some Observations

We have created this corpus from a very focused source, the sports news domain and have incorporated sentences, unique words and stemmed words. We extensively cleaned the data using a combination of filtering and stop word list — employing both manual and automated process. Every entry is cross-validated using at least two, and sometimes three annotators. We have manually corrected misspelling and stemming errors. So this is not just an annotated and ground truth corpus on sentiment analysis, it is also a corpus for training stemming engines.

It was also very important for us to build in auditability in the corpus. That is why every word and root word is cross-referenced against the source sentences. This way this corpus can be adopted for other NLP solutions with ease.

The other aspect of our corpus design is the transparency of the data collection process. It is a natural extension of the auditability of the data mentioned above.

VI. Open Source Licensing

Not-For-Profit and academic organizations and government agencies may use this corpus for noncommercial linguistic research and education only. For-profit organizations may use this corpus after signing a commercial technology development contract. Not-For-Profit and academic organizations and government agencies cannot use this corpus to develop or test products for commercialization, nor can they use this in any commercial product or for any commercial purpose.

VII. Conclusions and Future Work

We have presented a Bangla corpus specifically targeted for sentiment analysis. We described the methodology, source and clean up process. In the future, we plan to extend the corpus to support aspect based sentiment analysis for Bangla, where different clauses of a single sentence may have different sentiments. We also plan to extend this by adding sentiments for phrases, idioms and clauses. In addition, we plan to offer a set of machine learning models that can use this corpus.

References

[1] Adrija Roy and Abhishek Anand Singh. Sentimental Analysis (Bengali) https://github.com/abhie19/Sentiment-Analysis-Bangla-Language.

[2] English & Bengali Online Dictionary & Grammar. http://www.english-bangla.com/bntoen/index/%E0%A6%9C%E0%A6%9F%E0%A6%BF%E0%A6%B2.

[3] Tazim Hoque. Word and Doc2Vec file for Bengali Sentiment Analysis. https://www.kaggle.com/tazimhoque/bengali-sentiment-text/.

[4] Atik Rahman. Bangla Aspect Based Sentiment Analysis Dataset. https://github.com/AtikRahman/Bangla_ABSA_Datasets.

[5] Md. Atikur Rahman and Emon Kumar Dey. Datasets for Aspect-Based Sentiment Analysis in Bangla and its Baseline Evaluation. 4 May 2018. Institute of Information Technology, University of Dhaka, Dhaka 1000, Bangladesh. https://res.mdpi.com/data/data-03-00015/article_deploy/data-03-00015.pdf?filename=&attachment=1

[6] Asif Hassan, Mohammad Rashedul Amin, Abul Kalam Al Azada, Nabeel Mohammed. Sentiment Analysis on Bangla and Romanized Bangla Text (BRBT) using Deep Recurrent models. 24 Nov 2016. Dept. of Computer Science and Engineering University of Liberal Arts Bangladesh (ULAB), Bangladesh. https://docs.google.com/viewerng/viewer?url=http://resources.apurbatech.com/publication/upload/1610.00369.pdf

[7] D. Das and S. Bandyopadhyay, Developing Bengali WordNet Affect for Analyzing Emotion. Int. Conf. on the Computer. Processing of Oriental Languages, pp. 35-40, 2010.

[8] Cliff Goddard. Natural Language Processing, Edition: 2nd edition, Chapter: 5, Publisher: CRC Press, Taylor & Francis, Editors: Nitin Indurkhya, Fred J. Damerau, pp.92-120

[9] Mohammad Daniul Huq. Semantic values in Translating from English to Bangla, The Dhaka University Journal of Linguistics: Vol. 1 No.2 August, 2008 Pages: 45-66.

[10] Rafi Kamal. Bangla Stemmer. https://github.com/rafi-kamal/Bangla-Stemmer.

[11] Sazedul Islam. Rule based Bengali Stemmer written in python. https://pypi.org/project/py-bangla-stemmer/.