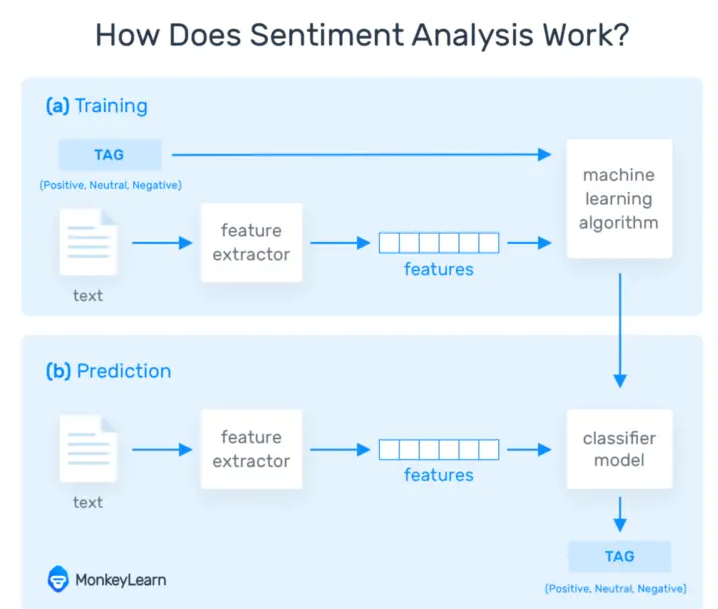
Resources on Sentiment Analysis

|  |  |  |  |
| --- | --- | --- | --- |
|  | Title & Link | Work Done | Date & Data Set |
|  | <https://www.researchgate.net/publication/335578758_An_Automated_System_of_Sentiment_Analysis_from_Bangla_Text_using_Supervised_Learning_Techniques> | This paper deals with six individual emotion classes- happy, sad, tender, excited, angry and scared. Here, we proposed two methods of machine learning techniques- Naïve Bayes Classification Algorithm and Topical approach to extract the emotion from any Bangla text. | Feb, 2019  used manually created data corpus with 7,500+ sentences as learning materials ☹ |
|  | <https://arxiv.org/ftp/arxiv/papers/1610/1610.00369.pdf>  Sentiment Analysis on Bangla and Romanized Bangla Text (BRBT) using Deep Recurrent models | a substantial textual dataset of both Bangla and Romanized Bangla texts have been provided…  this dataset have been tested in Deep Recurrent model… | 2016  prepared dataset of 9k+ texts, kept private, ☹ |
|  | Exploring Word Embedding For Bangla Sentiment Analysis  <https://ieeexplore.ieee.org/abstract/document/8554443/> (available on RG) | experimented current state of the art word embedding methods Word2vec SkipGram and Continuous Bag of Words with an addition Word to Index model for SA in Bangla language | 2018  created own dataset☹ |
|  | Design an Empirical Framework for Sentiment Analysis from Bangla Text using Machine Learning  <https://sci-hub.tw/10.1109/ECACE.2019.8679347> | quantifies total positivity and negativity against a document or sentence using Random Forest Classifier to classify sentiments | Feb 2019  ??? |
|  | N-Gram Based Sentiment Mining for Bangla Text Using Support Vector Machine  <https://sci-hub.tw/10.1109/icbslp.2018.8554716> | ……….. | Sep 2018 |
|  | Sentiment analysis for Bangla sentences using convolutional neural network  <https://sci-hub.tw/10.1109/ICCITECHN.2017.8281840> | used Adaptive Neuro-Fuzzy Inference System to predict the polarity of Bangla tweets and used fuzzy rules to represent semantic rules that are simple but greatly influence the actual polarity of the sentences | Dec 2017 |
|  | Sentiment Analysis Using Deep Learning Techniques: A Review  <https://pdfs.semanticscholar.org/8892/24a64a5bc5f9e965f418a63b6768f7164993.pdf> | highlights latest studies regarding the implementation of deep learning models such as deep neural networks, convolutional neural networks and many more for solving different problems of sentiment analysis such as sentiment classification, cross lingual problems, textual and visual analysis and product review analysis, etc. | 2017 |
|  | Sentiment analysis of bangla microblogs using adaptive neuro fuzzy system  <https://sci-hub.tw/10.1109/FSKD.2017.8393010> | done | July 2017 |
|  | Performing Sentiment Analysis in Bangla Microblog Posts  <https://sci-hub.tw/10.1109/ICIEV.2014.6850712> | …. | 2014 |
|  | Sentiment Analysis of Comments on Rohingya Movement with Support Vector Machine  <https://arxiv.org/abs/1803.08790> | used the result on a topic of interest | Mar 2018 |
|  |  |  |  |
|  | <https://github.com/abhie19/Sentiment-Analysis-Bangla-Language>  Dataset | **N-gram dataset of 4,600 sentences** | **2017** |
|  | <https://www.kaggle.com/tazimhoque/bengali-sentiment-text/download>  Dataset | **labeled dataset of 7,000 sentences** | **2018** |
|  | <https://www.mdpi.com/2306-5729/3/2/15/htm>  Publicly published dataset | **created one dataset from news comments, and another by translating, 5,000+**  [**https://github.com/AtikRahman/Bangla\_Datasets\_ABSA**](https://github.com/AtikRahman/Bangla_Datasets_ABSA) | **2018** |
|  | <https://blog.cambridgespark.com/50-free-machine-learning-datasets-sentiment-analysis-b9388f79c124> |  |  |
|  | <https://oxozle.com/awetop/keon-awesome-nlp/> | **awesome list of NLP tool, tutorials and datasets** |  |
|  | <http://amitavadas.com/SAIL/>  a challenge with dataset for Hindi and Bengali Sentiment Analysis | challenge with dataset for Hindi and Bengali Sentiment Analysis. Dataset for Bengali collected. (SentiWords) | 2015 |
|  | <https://github.com/krayush07/sentiment-semeval-absa> | huge dataset of SentiWords (erroneous) |  |
|  | <https://github.com/Kyubyong/wordvectors> | Bengali Word Vector built from a huge dataset |  |
|  |  |  |  |
|  | <https://cloud.google.com/natural-language/pricing> | **can be used for performance comparison \*\*\*\*\*** |  |
|  | <https://www.twilio.com/blog/2017/12/sentiment-analysis-scikit-learn.html> | For this post, we’ll use a pre-labeled dataset consisting of Twitter tweets that are already labeled as positive or negative. Using this data, we’ll build a model that categorizes any tweet as either positive or negative with [Scikit-learn](http://scikit-learn.org/). | 2017 |
|  | <http://www.laurentluce.com/posts/twitter-sentiment-analysis-using-python-and-nltk/> | This post describes the implementation of sentiment analysis of tweets using Python and the natural language toolkit [NLTK](http://www.nltk.org/). The post also describes the internals of NLTK related to this implementation. | 2012 |
|  | <https://monkeylearn.com/sentiment-analysis/>  **All you need to know on the topic and many other topics as well.** | Basics, how does it work, use cases, resources |  |
|  | <https://towardsdatascience.com/sentiment-analysis-concept-analysis-and-applications-6c94d6f58c17> | Good idea, analyzed Uber customer feedback from facebook and twitter Uber page. A good idea of cleaning data. |  |
|  | <https://data-flair.training/blogs/machine-learning-software/> | lots of lots of codes in many fields |  |
|  | <https://www.pyimagesearch.com/2019/10/21/keras-vs-tf-keras-whats-the-difference-in-tensorflow-2-0/?utm_source=facebook&utm_medium=ad-21-10-2019&utm_campaign=21+October+2019+BP+-+Traffic&utm_content=Default+name+-+Traffic+-+Image+1&fbid_campaign=6125643009046&fbid_adset=6125648754446&utm_adset=21+October+2019+BP+-+Email+List+-+Worldwide+-+18%2B&fbid_ad=6125648754846&fbclid=IwAR2WKeTYOMZeKeFnOMYSo_hCg5mxPFBCp1xrdlBkcIGr-uSFbXV95QZBhUM> | keras vs. tf2 vs. tf.keras |  |



|  |
| --- |
| Using SentiWordNet (ILs): Please cite the following references: |
|  |
| 1. A. Das and B. Gambäck. Sentimantics: The Conceptual Spaces for Lexical Sentiment Polarity Representation with Contextuality, In the 3rd Workshop on Computational Approaches to Subjectivity and Sentiment Analysis (WASSA), ACL 2012, Pages 38–46, Jeju, South Korea. [pdf](http://aclweb.org/anthology/W/W12/W12-3707.pdf) |
|  |
| 2. A. Das and S. Bandyopadhyay. Dr Sentiment Knows Everything! ACL/HLT 2011 Demo Session, Pages 50-55, June, Portland, Oregon, USA. [pdf](http://aclweb.org/anthology/P/P11/P11-4009.pdf) |
|  |
| 3. A. Das and S. Bandyopadhyay. SentiWordNet for Indian Languages, In the 8th Workshop on Asian Language Resources (ALR), COLING 2010, Pages 56-63, August, Beijing, China. [pdf](http://aclweb.org/anthology/W/W10/W10-3208.pdf) |

Asimuzzaman, … : There have been several researches on Sentiment Analysis using various techniques such as using word frequencies and TF-IDF [1] method etc. without considering the semantic structure of the sentences. However, there are grammatical patterns that influence the semantics of a sentence and simple frequencybased methods fail to incorporate the effects of such rules in their analysis.

We have used Bangla tweets for our training and testing data. We have collected the Bangla tweets using Twitter APIs, preprocessed it to remove neutral characters and words, and modified it into the form accepted by the Adaptive Neuro-Fuzzy Inference System (ANFIS).

identifying opinionated texts and determine their polarities

In paper [3], the authors used fuzzy sets to model semantic polarity by incorporating different semantic rules as rules of the fuzzy algorithm. Wordnet is an English lexical database that stores related nouns, verbs, adjectives and adverbs as similar groups of synonyms representing a similar concept. SentiWordNet is used to assign 3 scores on positivity, negativity and objectivity of words using WordNet. In paper [3], the authors assigned Parts of Speech (POS) tagging and polarity from SentiWordNet to each word in their opinion lexicon. The authors used the augmented lexicon to design semantic rules for their hybrid model.

precursors of this paper is by Chowdhury, S. & Chowdhury, W. [6]. Their work paved the way in understanding how sentiment analysis should be carried out in Bangla using semi-supervised approach to train their microblog corpus. The dataset has been collected through the Twitter API, filtered as such that it only extracted Bangla tweets. Since the extracted data was noisy it was preprocessed through tokenization, normalization and POS tagging. Tokens such as username, hashtags and URL links are cleared out of the datasets. The constructed lexicon is an association of positive / negative sentiments with words. We further take this approach and include sentiments like neutrality. The paper [6] is based on sentiment analysis on Bangla texts using SVM and Maximum Entropy. The authors collected from Bangla tweets using Twitter API v1.1, preprocessed and tokenized the data and then built the Bangla sentiment lexicon by first manually labeling the words and then checked the accuracy by translating them back to English and compared the results of the translated words using SentiWordNet.

To develop the fuzzy part of our paper, we got insights Dalal & Zaveri [7]. Linguistic hedges like modifiers (e.g.,“not”) concentrators (e.g., “very,” “extremely”), and dilators (e.g., “quite,” “almost,” and “nearly”) can improve the efficiency of sentiment classification. The authors incorporated fine-grained classification into multiple output classes like “very positive,” “positive,” “neutral,”

This paper [8] tackles Bangla language parsing using LFG (Lexical Functional Grammar) formalism. Extensive computational, linguistic, and psycholinguistic research has led to the evolution of Lexical Functional Grammar that provides a fundamental tool set of constructs for describing the common properties of all natural languages

In our research, we have used Bangla tweets, preprocessed to remove unwanted characters, performed POS tagging and then organized the data into separate training and testing datasets accepted by our model. For designing our model, we have ANFIS from Matlab’s Fuzzy Toolbox.



Sentence -> Tokens (NLTK) -> Translation (Samsad) -> POS Tagging (NLTK) -> Polarity value (SentiWordNet) -> Noun, Adj, Adv, V -> Train

Stemming [12]…

Abu Taher (N Gramm, SVM)…

Sentence -> Tokens (NLTK) -> Translation (Samsad) & POS Tagging (NLTK) -> Polarity value (SentiWordNet) wise selection of Noun and Verb; Adj, Adv -> Emoticons marked -> Unnecessary Punctuation removed -> Negativity Separation -> Verb Stemming -> Removal of unnecessary words -> N-gram selected -> SVM generated

\*\*\*\* We have vector represented database