
SENTIMENTAL ANALYSIS ON RESTURANT REVIEWS

Ananya Mukherjee
2018801009

SENTIMENTAL ANALYSIS

The process of computationally identifying and categorizing opinions expressed in a piece of text, especially in order to determine whether the writer's attitude towards a particular topic, product, etc. is positive, negative, or neutral.

FOR INDIAN LANGUAGE

- Sentimental analysis in Indian Language is a challenge due to limited resources available online.
- Scarcity of various tools & resources.
POS Tagger, chunker, Dependancy Parser,
Sentiment Lexicon (Pos/Neg words)
- Absence of benchmark datasets.
Quantity of reviews, Quality of reviews.

LEVELS OF ANALYSIS

- Document Level
- Sentence Level
- Phrase Level
- Aspect Level

DATASET

- Restaurant Reviews in Bengali is obtained.
- The data-set is human annotated.
- 2059 review comments is being annotated with category wise polarity.
- **Category** : Ambiance, Food, Service, Price, Anecdotes/Miscellaneous
- **Polarity** : Negative, Positive, Neutral, Conflict

Dataset Link:

https://github.com/AtikRahman/Bangla_ABSA_Datasets

DATASET

| Text | Category | Polarity |
|--|-----------|----------|
| শুধুমাত্র,রিডামিং ফ্যাক্টর খাদ্য ছিল,পুরোপুরি ন্যায্য হওয়ার | anecdotes | negative |
| খাবার একদমই ব্যতিক্রমী, একটি খুব সক্ষম রান্নাঘর যা গবে | food | positive |
| যেখানে গার্বিয়েলা লোকালি আপনাকে শুভেচ্ছা জানায়, এবং | service | positive |

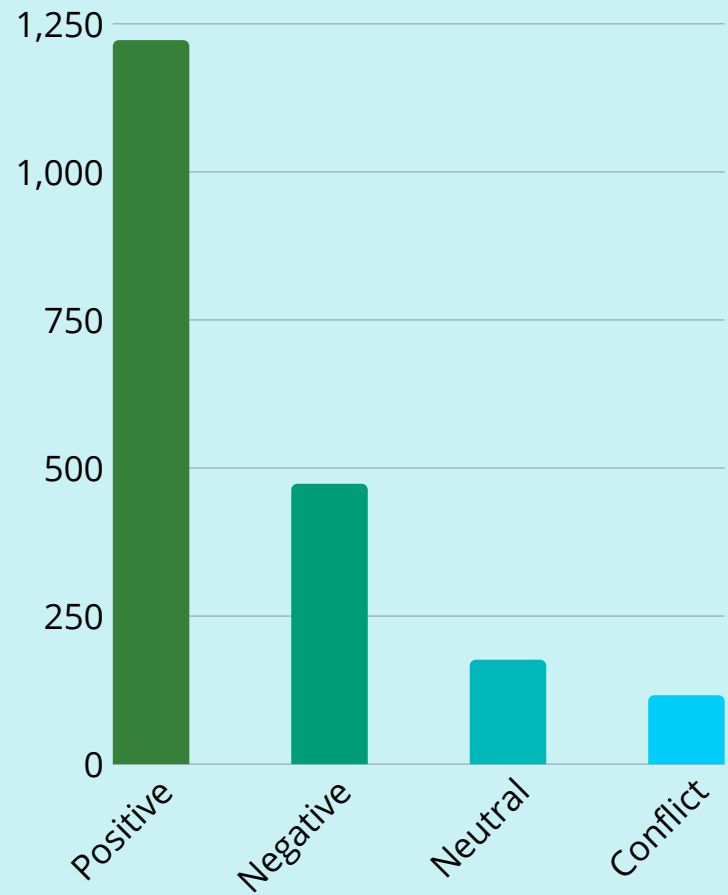
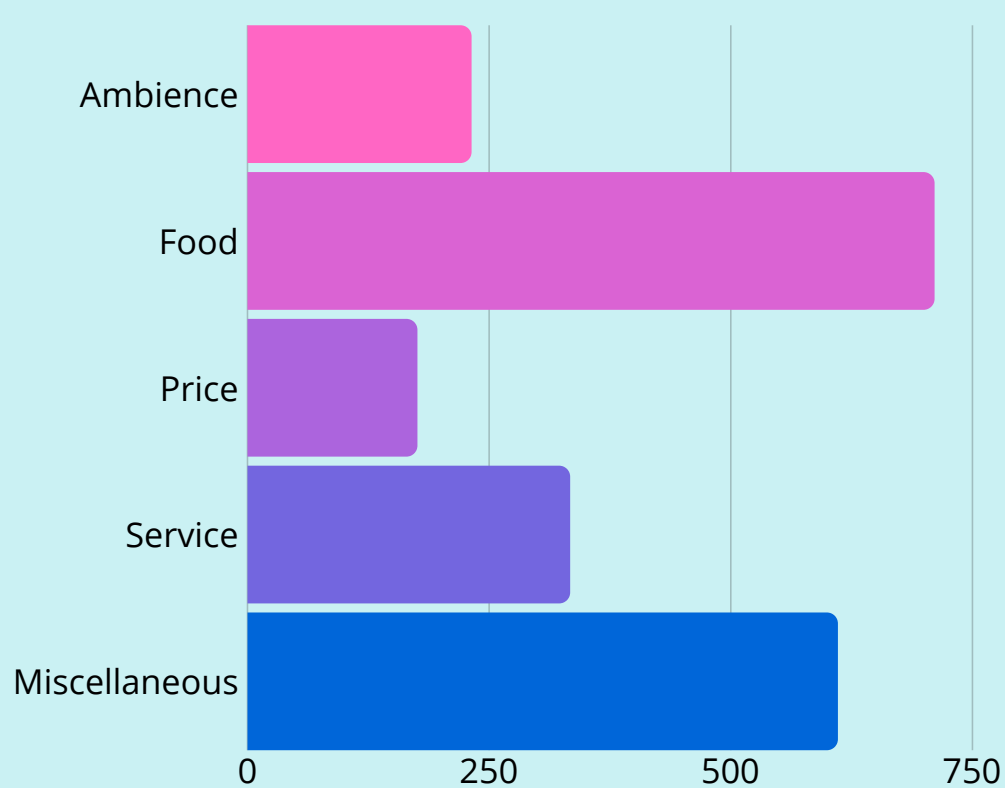
https://github.com/AtikRahman/Bangla_ABSA_Datasets

DATASET LABEL ADDITION

- Data-set has category wise annotation.
- Final label of the review is required.
- Label annotation was added by a Bengali Reader.

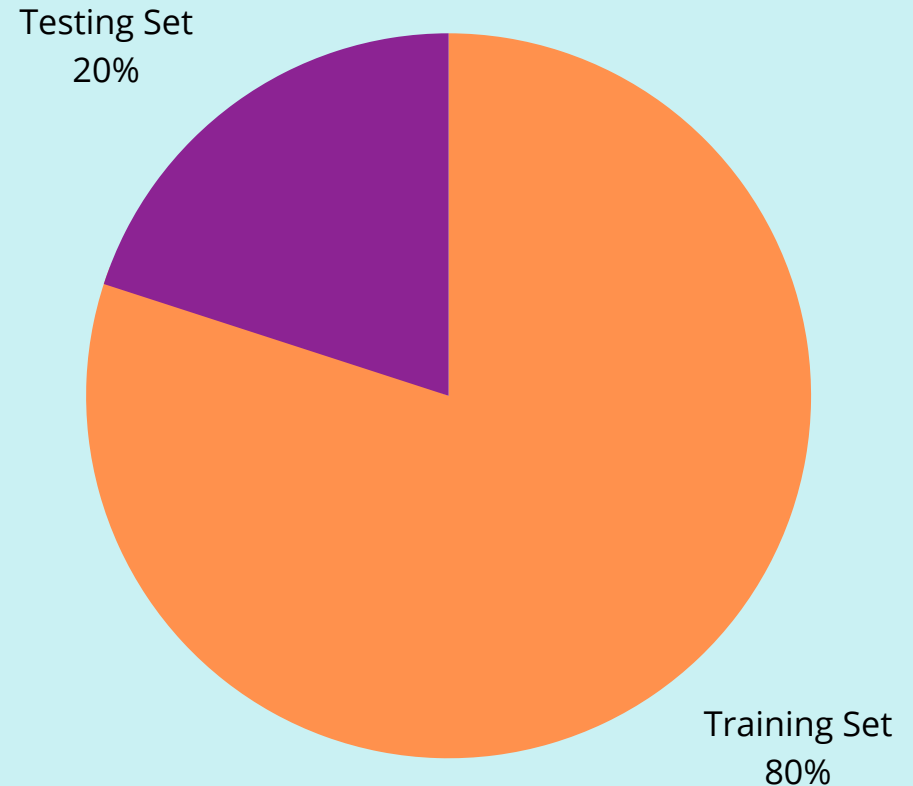
| Text | Category | Polarity | Label |
|---|----------|----------|-------|
| আমরা সবসময় মনোনিবেশ এবং প্রস্পট সেবা পেয়েছি,পরিবেশ খাঁটি এবং শিথিল হয়। | service | positive | 1 |
| খাঁটি, ফরাসি খাবার উল্লেখযোগ্য অংশ টাটকা,। | food | positive | 1 |
| তিনি উপর করা অভিনয় করেছেন,পানীয় আপ screwed,। | service | negative | -1 |
| এক নিরামিষ ভ্যানিটি (এবি এর ধন) আসলে একটি বেশ আশ্চর্য ছিল - এটি সুস্বাদু ছিল এবং ছিল শীতকালীন | food | positive | 1 |
| আপনি ভুল করতে পারেন না যে,\$ 5 এর জন্য 5 টি ডাম্পলিং। | price | positive | 1 |
| উপায় কিন্তু,ব্যয়বহুল পিজার | price | negative | -1 |
| এছাড়াও তরুণ কর্মীরা গরমে অনলস | service | positive | 1 |
| তারা জানেন যে স্ট্যাক অ পুইয়ের এবং ওগলেটের মত ফরাসি ক্লাসিকগুলি কীভাবে রান্না করা যায় না তা পুি | food | positive | 1 |
| এই মূল্য এবং খ্যাতি সঙ্গে একটি জায়গা থেকে আশা করি,সেবা হিসাবে শালীন ছিল কিন্তু মসৃণ নয়। | service | conflict | 0 |
| 45 এক মাসের NYC দেখার জন্য,আমরা নেপলসে একটি বিস্ময়কর খাবার ছিল | food | positive | 1 |
| ভাল পানীয় | food | positive | 1 |

DATASET DISTRIBUTION



DATASET SPLIT

- Training Set : 1647 records
- Testing Set : 412 records



APPROACH

- Each comment is cleaned, tokenized into words and fed to the NLTK POS tagger.
- All the adjectives and nouns are considered.
- Five clusters are formed for each category (service, price, food, ambiance, misc).
- Each category having 3 subsets (Positive, Negative, Neutral)
- Calculate the polarity wise probability for each category.
$$\text{NegativeProbability} = \text{NegativeWords} / \text{Total}$$
$$\text{PositiveProbability} = \text{PositiveWords} / \text{Total}$$
$$\text{NeutralProbability} = \text{NeutralWords} / \text{Total}$$
- Calculated overall Positive, Negative & Neutral probabilities.

APPROACH

- For every new test set review, find the number of words matched in each category for positive, negative & neutral.
- Find the likelihood of the review being positive/negative/neutral.
- **PositiveLikelihood** = $\text{MatchWords}(\text{pos})(5 \text{ Ctg}) / \text{TotalWords} * \text{CategoryProbability}(\text{pos})$
- Higher likelihood probability is the winner.
- Tried to formulate it as a Naive Bayes Classification, assuming the features to be independent.

RESULTS

| Text | Category | Polarity | Actual | Predicted |
|--|-------------|----------|----------|-----------|
| তিনি উপর করা অভিনয় করেছেন,পানীয় আপ screwed,I | service | negative | negative | positive |
| এক নিরামিষ ভ্যানিটি (এবি এর ধন) আসলে একটি বেশ আশ | food | positive | positive | positive |
| আপনি ভুল করতে পারেন না যে,\$ 5 এর জন্য 5 টি ডাম্পলিং | price | positive | positive | positive |
| উপায় কিন্তু,ব্যয়বহুল পিজার | price | negative | negative | negative |
| এছাড়াও তরুণ কর্মীরা গরমে অনলস | service | positive | positive | positive |
| মানুষ প্রচুর,মহান Vibe, | food | conflict | neutral | negative |
| উপায় কিন্তু,ব্যয়বহুল পিজার | price | negative | negative | negative |
| যথেষ্ট বলেছ | anecdotes/n | neutral | neutral | negative |
| প্রস্তাবিত। | anecdotes/n | positive | positive | negative |
| ভয়ানক অভিজ্ঞতা | anecdotes/n | negative | negative | negative |

PERFORMANCE

Accuracy : 57.52 %

Confusion Matrix :

| Negative | Neutral | Positive |
|----------|---------|----------|
| 2 | 0 | 77 |
| 4 | 0 | 86 |
| 8 | 0 | 235 |

Classification Performance:

| | precision | recall | f1-score | support |
|-----------|-----------|--------|----------|---------|
| | | | | |
| negative | 0.14 | 0.03 | 0.04 | 79 |
| neutral | 0 | 0 | 0 | 90 |
| positive | 0.59 | 0.97 | 0.73 | 243 |
| | | | | |
| avg/total | 0.38 | 0.58 | 0.44 | 412 |

CONCLUSIONS

- Data collected has been updated by adding a new label column.
- 5 Categorical Clusters are formed based on annotations
- Probabilities has been calculated polarity wise for each category.
- Using Naive Assumption, likelihood is calculated for each review in the test set.
- Final label prediction is done based on the maximum likelihood.

FUTURE WORK

- Can be extended to other Indian Languages provided the POS tagger is precise.
- Collect more data, to have a balance distribution of sentiments rather than being biased.

THANK YOU!
