1. Model Training
2. Feature Engineering

In feature engineering we convert the text data into numeric form.

There are three methods:

1. Bag of Word
2. TF-IDF
3. Word 2 vec
4. Remove space, newline and blank line.
5. Contraction mapping (Text Expand)
6. Handling the accented character.
7. Cleaning:
8. Tokenization
9. Remove punctuations
10. Remove stopwords
11. Remove word whose length less than 2
12. Normalizations
13. Stemming and Lemmatizations

There are three methods for EDA

1. Ngram: unigram, bigram and trigram
2. Word Cloud
3. Key phrase Extraction
4. YAKE
5. RAKE

The review detects in multiple language. firstly, convert all review in single language (“English”)

1. Initial Data Preprocessing

Train the model:

Naïve Bayes

To build a model that will determine the tones (positive, negative and neutral) of the text or the review. The resulting model will have to classify the positive review, negative review and neutral review.

Data Available in CSV format

In data we need only two feature dependent feature (stars) and independent feature (review)

1. Data Preprocessing
2. Exploratory Data Analysis
3. Data Gathering
4. Problem Statement

**Sentiment Analysis**