Sentiment Analysis of Phones sold on Amazon

Abstract

Reviews are being used to gain insight into consumer related decisions as the understanding of its associated sentiment either positive, negative, or neutral provides organizations very important market awareness and the ability to proactively address issues early. This paper analyzes the sentiment given the review tone/text and predict the ratings based on the review and their correlation by using Natural language processing techniques and procedures using

Defining the Problem

With e-commers and social media public participation and producing content on products is relevant in most business to tap on those sentiments and make decisions on enhancing the public opinion on products and services they provide. And given the number of websites, social media platforms it is humanly not possible to employ humans on such a waste amount of data to discover patterns and specially the sentiment of such views from large volumes of data that could wary from a simple one liner to large blogs of text, different tones, vocabulary, slangs and local conditions manipulate the view of the consumer/reviewer to have a consistence view given the product/services are spread across the globe. In this analysis the data is coming from Amazon phone sales reviews.

Defining the target variable:

Given the review in plain text, the task is to find out if the reviewer has a positive, negative, neutral sentiment on that product/service along with uncover some of the common patters of words used for each brand of phones. Also, predict the rating based on the review tone/text.

Data Understanding:

Data set consist of 413840 samples of data with The name of the Product, Name of the parent, Price of the product, Rating of the product, Description of the user experience & Number of people voted the review.

Data was acquired in December 2016 by the web crawlers build to deliver data extraction service https://www.promptcloud.com/

Data source file acquired for this analysis is from https://data.world/promptcloud/amazon-mobile-phone-reviews

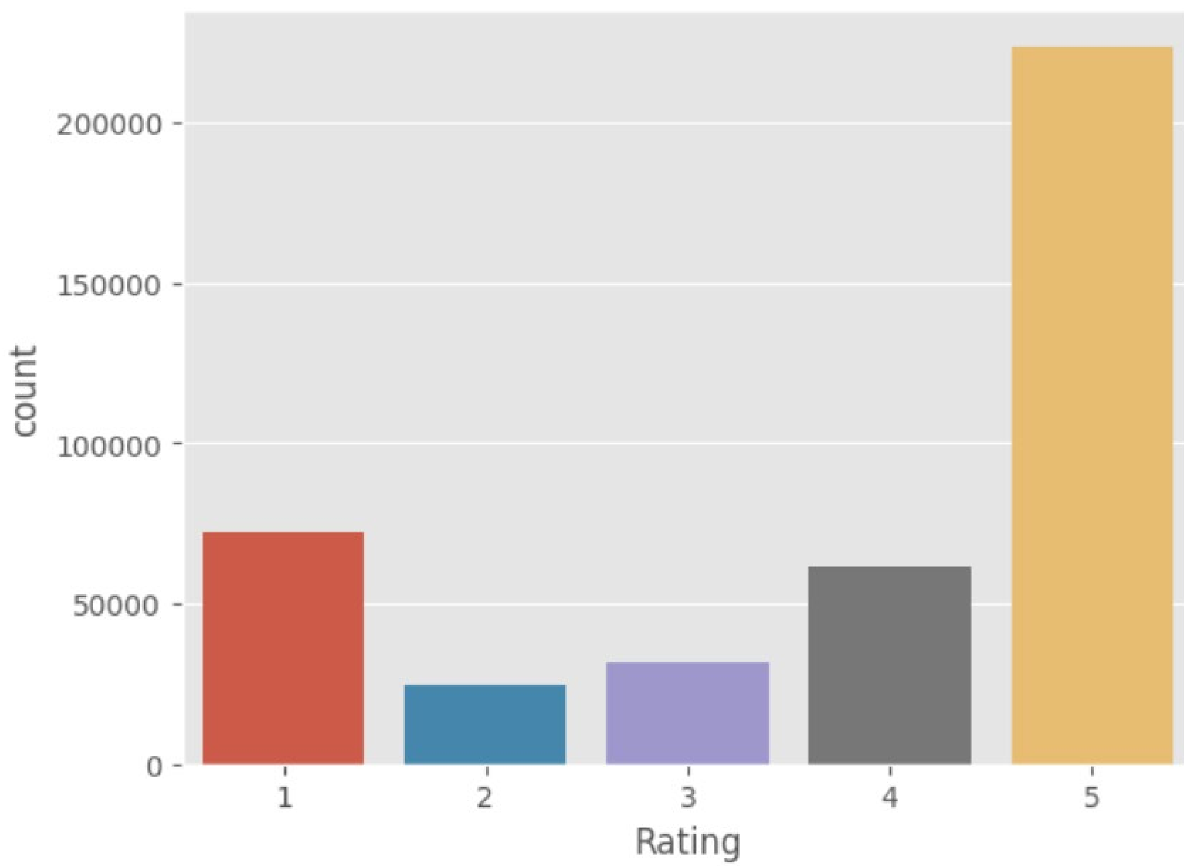
```
df.dtypes
```

```
Product Name      object
Brand Name        object
Price            float64
Rating             int64
Reviews           object
Review Votes     float64
dtype: object
```

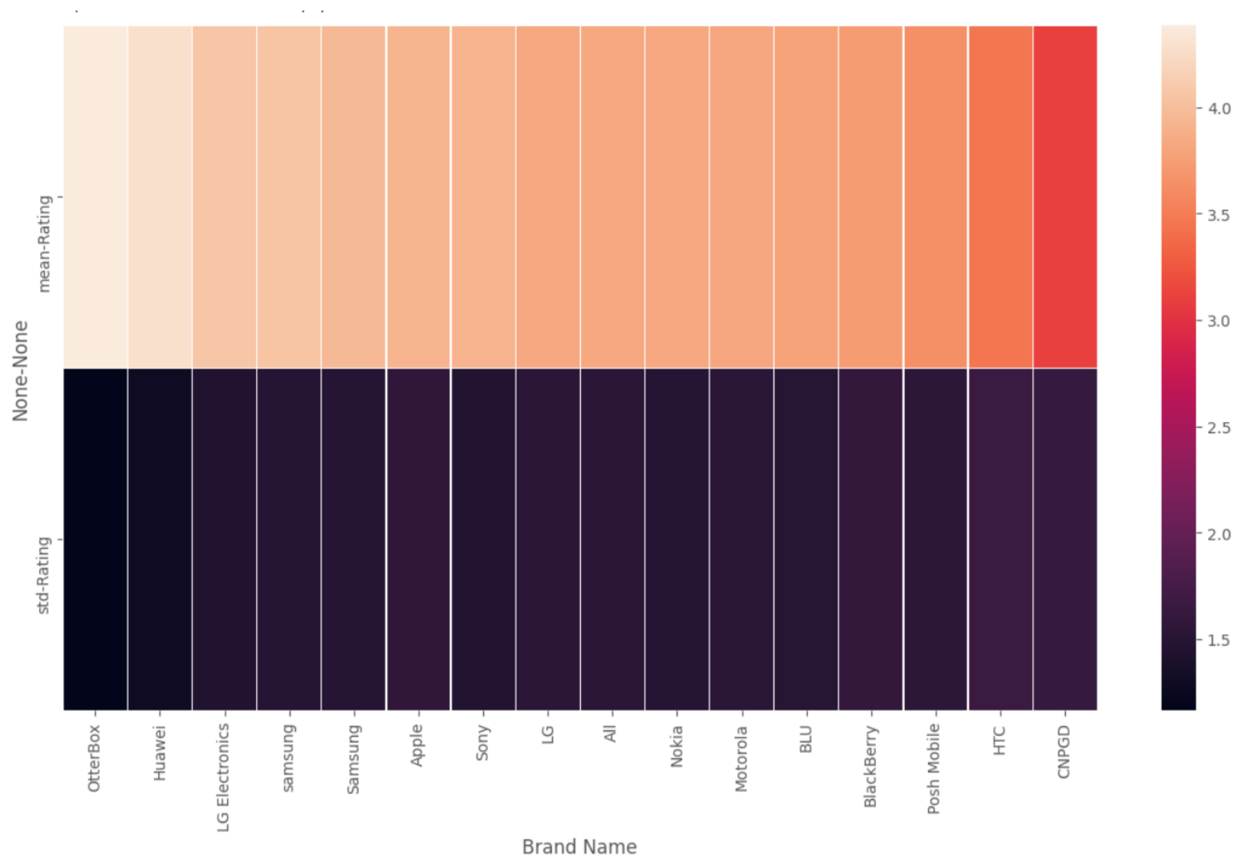
Data has large amounts of reviews by single reviewer and other brands of phones with very large reviews.

Data Exploration:

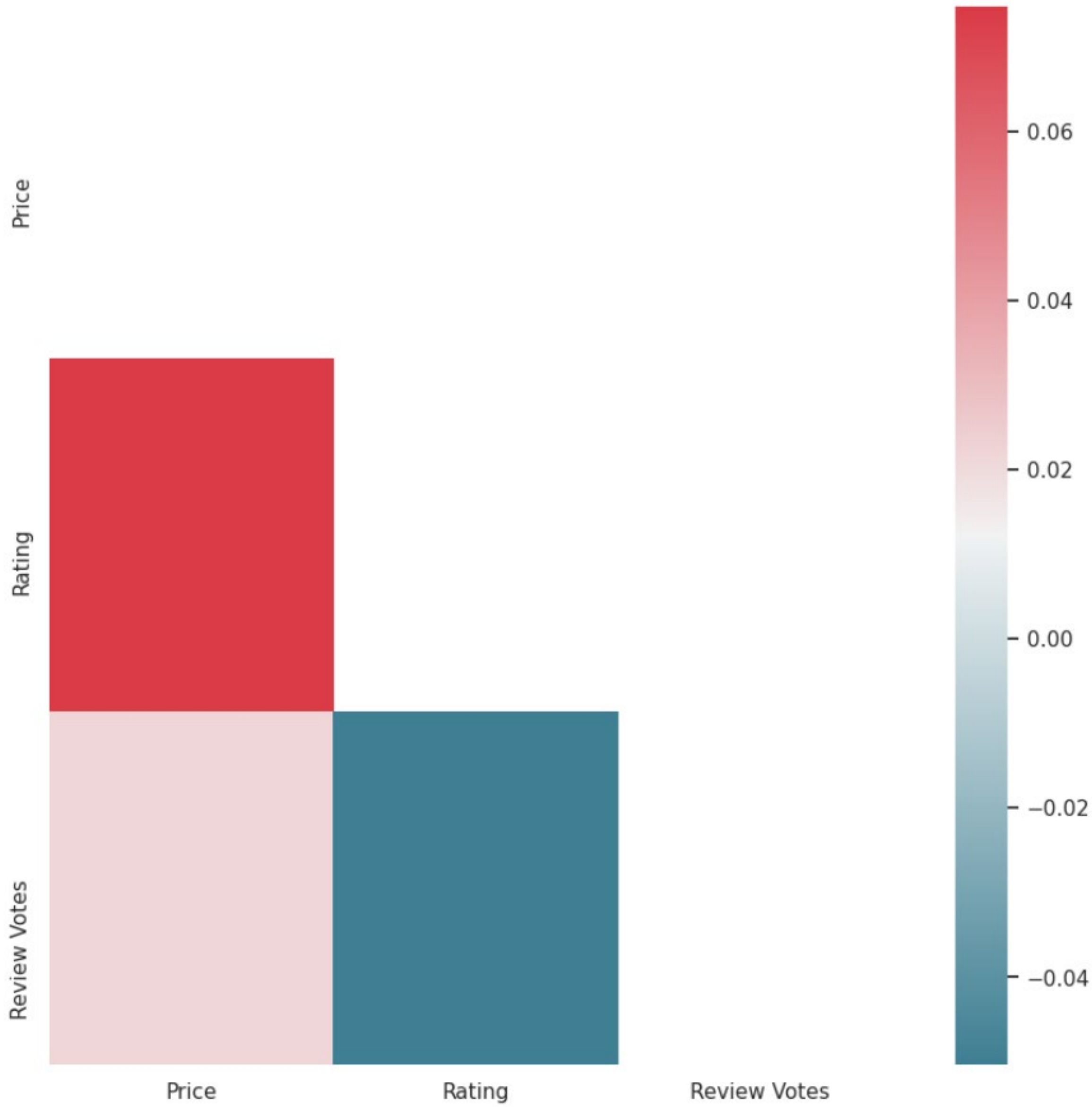
Performing EDA on the data set, the following observation are made against the rating given the number of reviews are made from the total sample of 413840, and it explains there more 5star ratings brand phones. It also concludes reviewers either they are extremely happy with the phones or completely dissatisfied with their purchase.



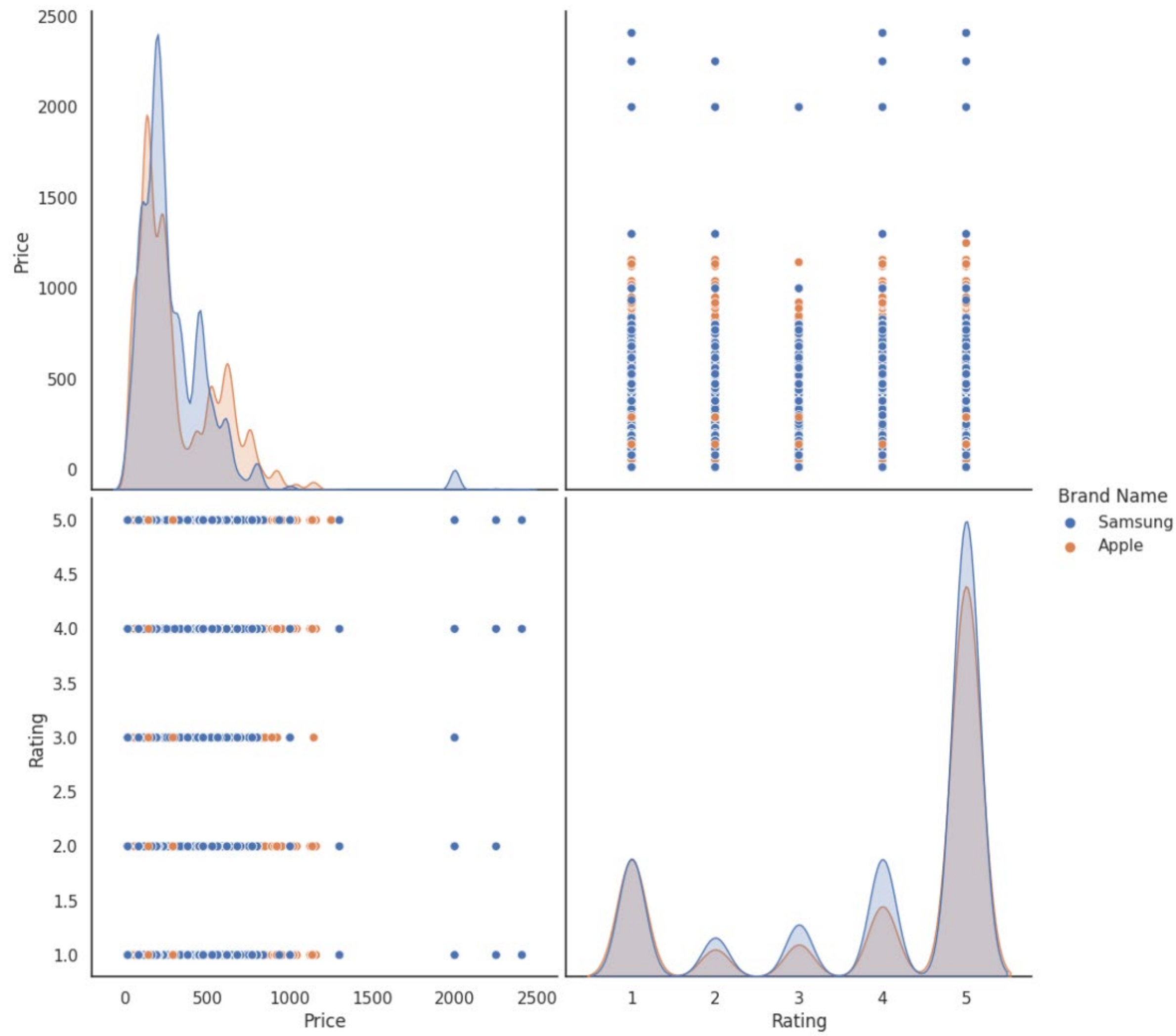
Since the data has quite a bit of brands, took top 15 brands by review counts and ploted the mean and std rating to depict the color scale. It is evident that these OtterBox has the highest mean ratings along with std ratings. Also most major phone brands fall in to the top 15 review counts and ratings.



Also did a co-relation heat map analysis and found that price and ratings are comparatively stronger correlation followed by price and review votes. It also indicates people spending higher prices tend to review the product and also the satisfaction on such products is expressed higher than the low-price products.



Comparing two popular brands price and ratings, it is evident that Samsung has more products at the lower end of the spectrum compared to Apple brand.



Data Preparation, transformations & Imputation:

Dropped n/a values and considered only the ratings greater than 3 to 1 and others to 0 and cons olidate the rating scenarios so to not having to deal with 5 different scenarios. And for sentiment analysis ignored 3 assuming they are neutral in nature. Used Stop words to remove stop words, punctuation, and used re library to normalize all texts removing non textual content from the rev iews and make those reviews in to tokens for use of predicting the words used for use in predicti ng the ratings. In the process also used lemmatization, grouping of words to be used as one singl e item (which is derived from Wikipedia).

Data Standardization & transformations:

Using bag of words model, used feature extraction methods to get count of vectors and fit the values through the transformation with the resulting text of review. Further fed to TF-ID to re-flect how much importance each word is to a situation of the review in general.

Modeling:

For sentiment analysis used Polarity score by passing the review text and got hold of each review sentiment value for each of positive, negative, and neutral value. For example a review got

```
'neg': 0.13, 'neu': 0.87, 'pos': 0.0, 'compound': -0.69, 'brand': 'Jethro'
```

On evaluating all reviews and got a mean of each positive , negative , neutral values for each brand and by specifically showed the top 15 brands sorted by highest mean values along with overall phones being sold on Amazon.

Brands that have the highest negative mean values (top 15)

```
neg_sentiment_top15[neg_sentiment_top15["min"] != neg_sentiment_top15["max"]]
```

Output Visualize

	mean	min	max
brand			
nokia	0.47950000000000004	0.323	0.636
Nuu Mobile	0.356	0.097	0.615
KROO	0.339	0.0	0.678
Ipro	0.273	0.0	0.546
Wool	0.262	0.0	0.524
Jiayu	0.2555	0.043	0.468
Higoo	0.214	0.0	0.608
Various	0.204	0.0	1.0
ifcane	0.20059999999999997	0.0	0.37
Eachbid	0.1855	0.0	1.0

Brands with highest Positive mean values are (top 15)

```
pos_sentiment_top15[pos_sentiment_top15["min"] != pos_sentiment_top15["max"]]
```

Output Visualize

	mean	min	max
brand			
Spicy World	0.7675000000000001	0.727	0.808
SGH-T199	0.7445999999999999	0.592	1.0
Saliency	0.63	0.516	0.744
LG Electronic	0.6105	0.388	0.833
JINHAIHUAHUI	0.603	0.402	0.804
Meizu	0.5836666666666667	0.238	0.799
Doogee	0.5654374999999999	0.0	1.0
Storm	0.5635	0.127	1.0
DIKOO	0.5532857142857143	0.0	1.0
iNew	0.5259999999999999	0.296	0.756

And for all the brands sold on Amazon, the mean values for all three sentiments are

```
Overall Negative mean 0.05441774840518049
Overall Negative median 0.0

Overall Neutral mean 0.6568210371158152
Overall Neutral median 0.731

Overall Positive mean 0.28805419969073154
Overall Positive median 0.19

Overall Compound mean 0.3607328977382892
Overall Compound median 0.5023
```

On evaluating the overall scores of sentiments by taking mean values, the general mode is neutral with 64%, while positive is 30% and rest negative.

On the second task of predicting the ratings based on reviews, passed the reviews through logistic regression, naïve bayes & randomforest classifier

logistic regression:-

```
Accuracy for Logistic Regression: 0.963273693441357
confusion matrix for Logistic Regression:
 [[17786  1679]
 [ 1127 55811]]
F1 score for Logistic Regression: 0.9754780298528333
Precision score for Logistic Regression : 0.970794920855801
recall score for Logistic Regression : 0.9802065404475043
AUC:  0.9469745776987073
```

naïve bayes :-

```
Accuracy for Naive Bayes Classifier: 0.93337957933589
confusion matrix for Naive Bayes Classifier:
 [[17047  2418]
 [ 2672 54266]]
F1 score for Logistic Regression: 0.9552023375754695
Precision score for Logistic Regression : 0.957342459953426
recall score for Logistic Regression : 0.9530717622677298
AUC:  0.9144243989864208
```

Random forest classifier

```
Accuracy for Random Forest Classifier: 0.9767941049435231
confusion matrix for Random Forest Classifier:
 [[18245  1220]
 [  553 56385]]
```

Based on these three models, Random Forest classifier **with 98% accuracy** and True positive are higher than other models and false positives are lowest among other models.

Conclusion:

Generalizing the review dataset and understanding the final outputs sentiment values of ratings and make normalizations among those rating reviews is the main assumption. Given the amount i.e 65% of neutral mean values we got in sentiment analysis and the number of neutral ratings (value 3) among the overall ratings, it is very helpful to business to take these values and strive towards improving the ratings to positive. Also, in predicting the rating values based on text values among blogs and other social media sites, that don't have necessarily a rating, the random forest classifier would be a good start to find mean ratings among the users.

Next steps: - Explore other models using LSTM and SVM as most papers suggest they might be faster and accurate given the products and services are used across the globe and nullify factors which be biased using random forest classifiers.

Code files are uploaded to GitHub repo at the following location.

https://github.com/11leven/portfolio/tree/gh-pages/DSC%20680

Q & A

1. Can this data set use for predicting $ amount of phones?

   Given the data points with $ amounts I think that would be a good use case to predict the brand phones given the ratings on the brands and products.

2. What kind of vectorization methods used?

   TfidfVectorizer used to gets tokenize documents, learn the vocabulary and inverse document frequency weightings

3. What are the issues with data?

   There are about 8 kinds of phones sold by Samsung for example and bringing them all one umbrella was a challenge since there are phones and accessories sold under phones.

4. What other use cases can be implemented on this review data?

   Recommendation engine would be a good use case to implement, since accessories are also part of the data set

5. What kind of sentiment scoring techniques used?

   SentimentAnalyzer Opinion mining to derive the opinion or attitude of a speaker

6. What are the challenges in the analysis?

   Very little features and solely relying on the review text and not much relation exists between the products vs brands feature

7. What are the run times of the sentiment analysis?

   Approximately 15min for 400k reviews on a 4core windows machine

8. Stopwords and Punctuations used for sentiment analysis?

   Used all parts of the text to process and identify the positive, negative, and neutral score per each review text.

9. What model evaluation techniques used for all models?

   Since this is a sentiment determination, there are no evaluations methods used. Used F1, precision and recall to evaluation techniques when predicting the review score.

10. Next steps?

    Build more LSTM and keras models to evaluate the sentiment and performance of models.

Acknowledgements & References:

1. Leveraging Deep Learning for Multilingual Sentiment Analysis https://www.linkedin.com/pulse/leveraging-deep-learning-multilingual-sentiment-parsa-ghaffari/
2. Training Deep Convolutional Neural Network for Twitter Sentiment Classification http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.703.6858&rep=rep1&type=pdf
3. Twitter Sentiment Classification using Distant Supervision https://www-cs.stanford.edu/people/alecmgo/papers/TwitterDistantSupervision09.pdf
4. Sentiment Classification using Machine Learning Techniques https://dl.acm.org/doi/pdf/10.3115/1118693.1118704
5. Thumbs Up or Thumbs Down? Semantic Orientation Applied to Unsupervised Classification of Reviews https://dl.acm.org/doi/pdf/10.3115/1073083.1073153
6. Learning subjective adjectives from corpora https://www.aaai.org/Papers/AAAI/2000/AAAI00-113.pdf
7. A Stochastic Parts Program and Noun Phrase Parser for Unrestricted Text https://aclanthology.org/A88-1019.pdf
8. Smokey: Automatic recognition of hostile messages https://www.aaai.org/Papers/IAAI/1997/IAAI97-209.pdf
9. Using Natural Language Processing to Power Chatbots https://discover.bot/bot-talk/behind-the-scenes-using-nlp-to-power-chatbot/
10. What Is a Chatbot? Everything You Need to Know https://www.ultimate.ai/blog/ultimate-knowledge/what-is-a-chatbot