# Amazon Mobile Phone Reviews Sentiment Analysis

Ghazal AYOBI

# Abstract

This project utilizes Amazon mobile phone reviews dataset from Kaggle, the link for the dataset can be found here. The dataset contains two tables; items and reviews. This project aims to answer how to use different sentiment analyzer on mobile phone reviews and evaluate the most common issues? The methods used in this project are natural language processing to analyze reviews. As a result of using different sentiment analyzer this project chooses Textblob sentiment Analyzer and finds that the most common problem in the mobile phone reviews is *Battery life.*

# Motivation

Customer reviews provide a wide range of information for decision making while shopping. When it comes to mobile phones, they are one of the vital devices to the human beings in this time. An individual buys a mobile phone and on average uses it for two to three years. Thus, decision making should involve understanding the opinion of the ones who already used a smartphone. Thus, the motivation behind this project is to understand the sentiment of mobile phone reviews and find out the most common issues related to customer review.

# Dataset(s)

This dataset for this project is called Amazon Mobile Phone Review which is taken from Kaggle. This link to the dataset can be found here. The The dataset contains two tables; Review and Items. The Items dataset contains information on each mobile phone and the Review table is the customers comments for all available smartphones. In order to access the data fast, I uploaded it on my GitHub repository.  The Items table contains information for 720 different types of mobile phones. The review table contains information for around 68000 customer reviews.

# Data Preparation and Cleaning

The original dataset contained two tables. Initially I merged the both data tables. Before beginning sentiment analysis. I explored the dataset and created necessary columns such as date and others. The main goal of this project is to analyze the reviews. Thus, natural language contains special characters, review are provided in various languages and other issues such as punctuation. In the first step I removed the punctuations and emojis. Secondly to narrow down the goal of project I only concentrate on the English language reviews. Thus, using language detection algorithm I only keep English language comments. The following step to have the comments text cleaned and remove stopwords. Thus, I tokenized the reviews. As in the natural language a word can have different forms. I used Lemmatization algorithm to get the base form of the words . As a result of cleaned text I used two sentiment analysis algorithms to evaluate comments sentiment. The most common project in the procedure was code run time. The dataset is quite time consuming thus, I used random samples of the data to run the code. The other problem I came across that the algorithms identify a review as Neutral however it is either positive or negative.

# Research Question(s)

This project aims to answer how to use different sentiment analyzer on mobile phone reviews and evaluate the best performing sentiment analyzer to find the most common issues?

# Methods

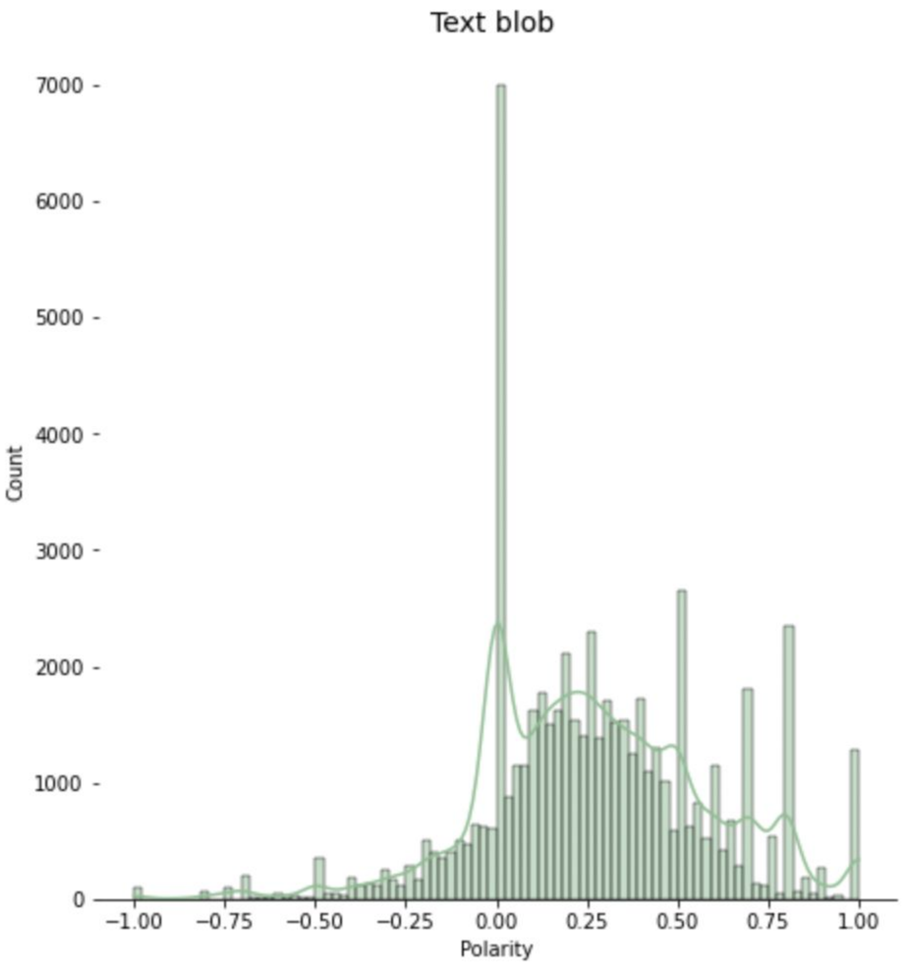
In the initial step, the dataset contains information about reviews text and customers ratings. Customers ratings provide a general perspective of a customer review. The most important decision in this project is I added a new column to classify reviews based on the customer rating. For example rating has the values from 1 to 5 which is highest. Thus I classify rating 4, and 5 as positive, 3 Neutral and the rest Negative.

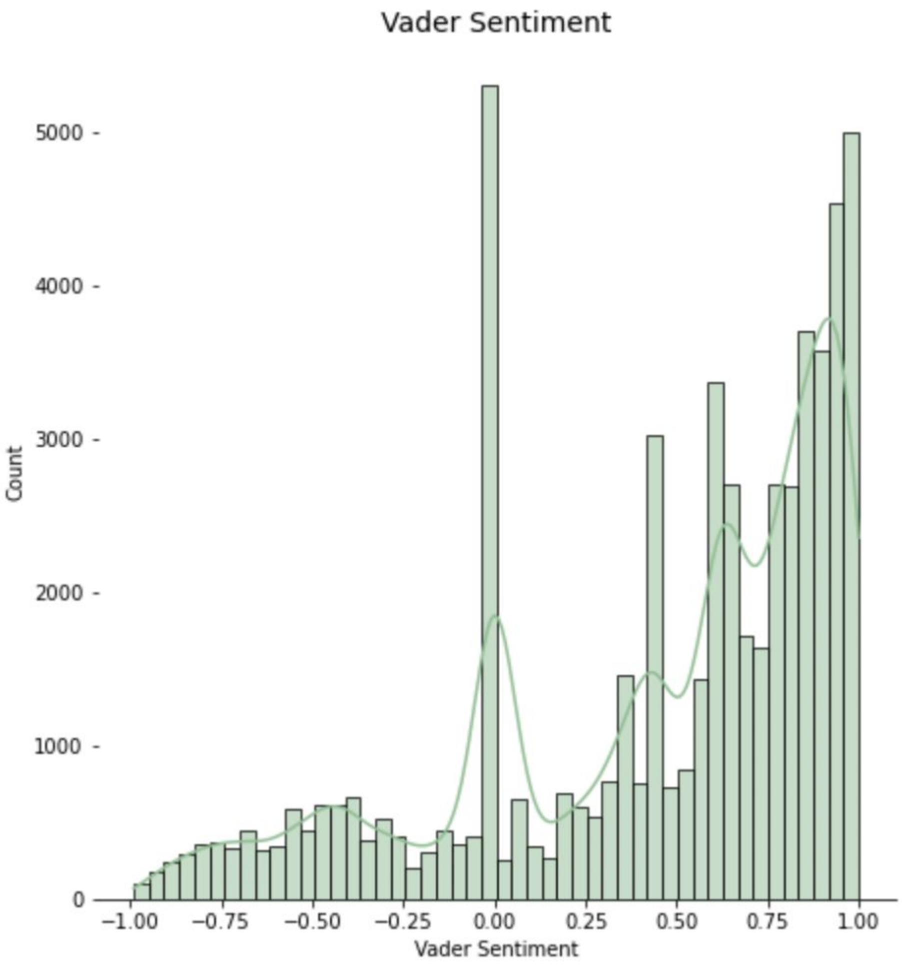
To clean the text, I created functions to remove punctuations, and emojis. I used language detection algorithm to keep english reviews and tokenized the words. As a result I applied two sentiment analyzer; Textblob and Vader and compared the result  based on the rating classifier created in the beginning. As a final step I used Naive Bayes classifier to select the best sentiment analyzer.  Based on the Naive Bayes classifier Textblob sentiment analyzer provides the best accuracy identifying the Positive, Negative and Neutral comments based test and train classification method.

# Findings

This project aims to analyze the Amazon Mobile Phone reviews sentiment. After initial data exploration. This project uses different sentiment analyzer to evaluate reviews. In the the following Findings slides we can see the sentiment score distribution of both the two main sentiment analyzer for this project.

In the following slide it is evident that Textblob sentiment analyzer has a close to normal distribution sentiment score. However, Vader sentiment analyzer has more positive sentiments


Text blob
7000
6000
5000
4000
3000
2000
1000
0
Count
−1.00
−0.75
−0.50
−0.25
0.00
0.25
0.50
0.75
1.00
Polarity


Vader Sentiment
5000
4000
3000
2000
1000
0
Count
−1.00
−0.75
−0.50
−0.25
0.00
0.25
0.50
0.75
1.00
Vader Sentiment

# Findings : Classification Report

The following Findings slide shows classification report to get Naive Bayes classifier report. The following classification reports show the precision, recall, f1-score, and accuracy score for each model which are text blob and Vader sentiment.

Text blob model indicates an accuracy of 0.78. The precision shows that the ratio of true positive to the sum of true and false positives, for negative sentiment is 1.00, for neutral is 0.99 and for the positive sentiment is 0.77 which is the highest among all models.

We can see that Vader sentiment shows an accuracy of 0.68 which is the lowest among all models. the precision of negative sentiment is 1.00,neutral is 0.78 and positive precision is 0.67 which is the lowest among other models.

# Finding : Classification Report

Textblob

| | precision | recall | f1-score | support |
|---|---|---|---|---|
| Negative | 1.00 | 0.01 | 0.02 | 7129 |
| Neutral | 0.97 | 0.01 | 0.02 | 6296 |
| Positive | 0.77 | 1.00 | 0.87 | 44841 |
| accuracy | | | 0.77 | 58266 |
| macro avg | 0.91 | 0.34 | 0.30 | 58266 |
| weighted avg | 0.82 | 0.77 | 0.67 | 58266 |

Vader Sentiment

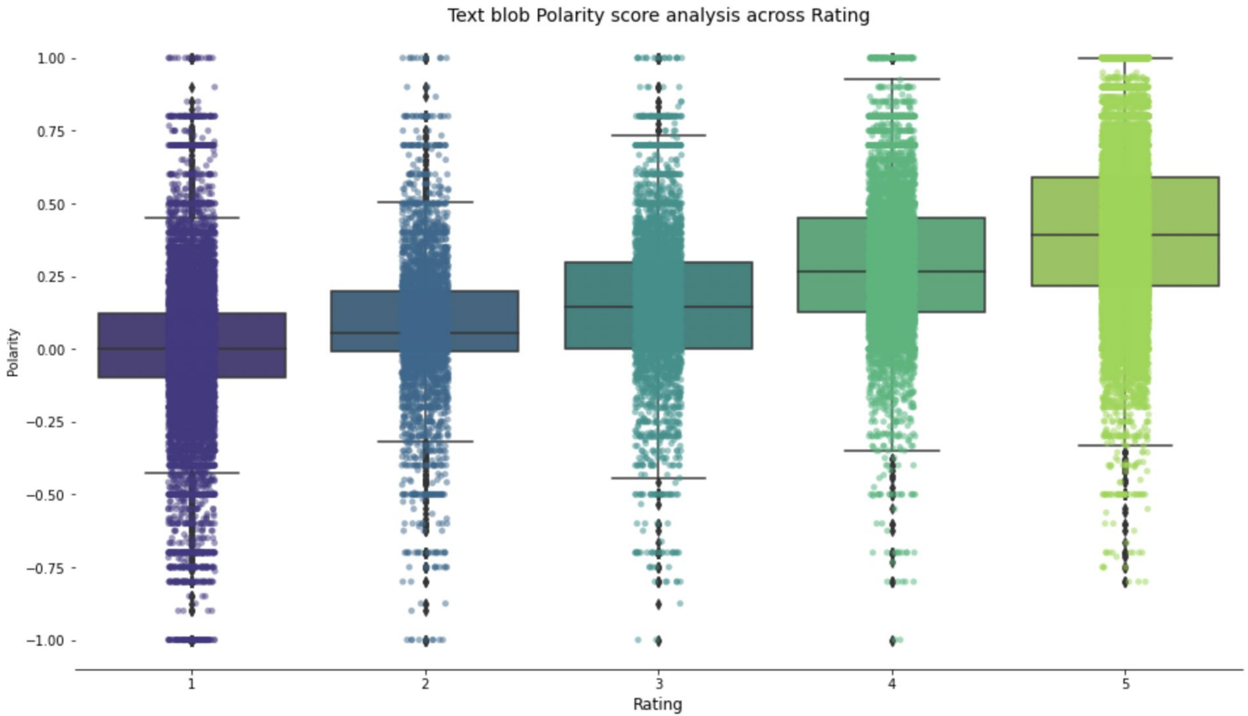
| | precision | recall | f1-score | support |
|---|---|---|---|---|
| Negative | 1.00 | 0.00 | 0.00 | 3828 |
| Neutral | 0.78 | 0.29 | 0.42 | 20501 |
| Positive | 0.67 | 1.00 | 0.80 | 33937 |
| accuracy | | | 0.68 | 58266 |
| macro avg | 0.82 | 0.43 | 0.41 | 58266 |
| ighted avg | 0.73 | 0.68 | 0.61 | 58266 |

# Findings : Textblob Test - Train Result

| | precision | recall | f1-score | support |
|---|---|---|---|---|
| Negative | 1.00 | 0.00 | 0.00 | 2178 |
| Neutral | 1.00 | 0.00 | 0.01 | 1877 |
| Positive | 0.77 | 1.00 | 0.87 | 13425 |
| accuracy | | | 0.77 | 17480 |
| macro avg | 0.92 | 0.34 | 0.29 | 17480 |
| weighted avg | 0.82 | 0.77 | 0.67 | 17480 |

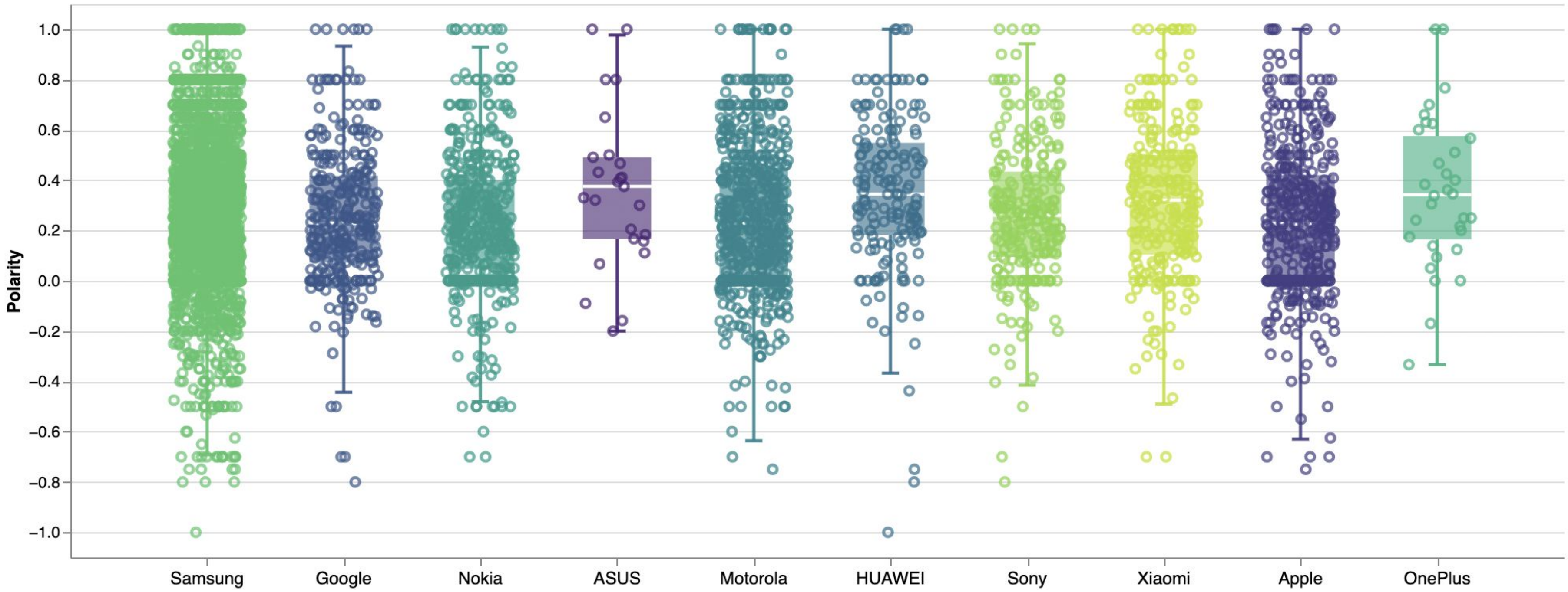
The result show classification report for Textblob, I divided the dataset into test and train sets. where the the test size is 30% of the observations and data is selected randomly. The accuracy of the model is 0.77 and negative and neutral has the precision of 1.00 and positive sentiment has precision of 0.77.

# Findings

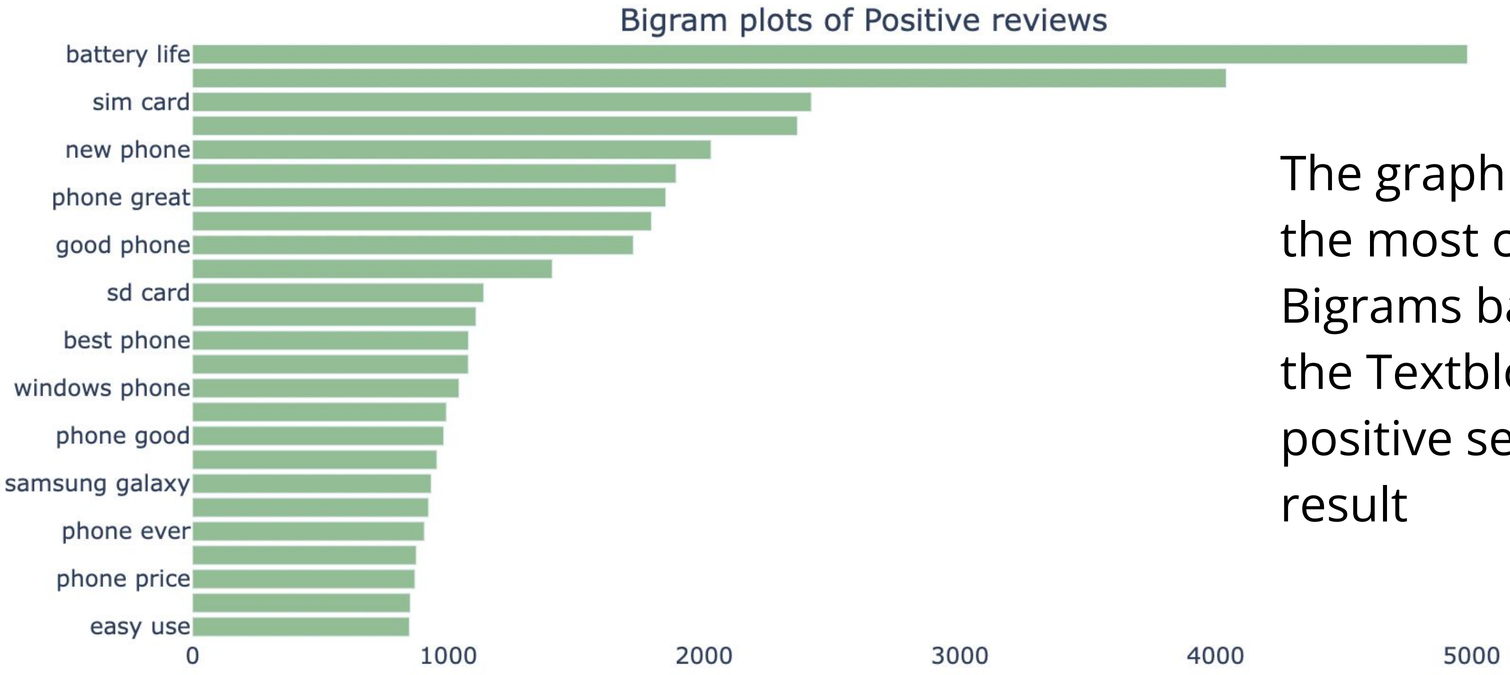



The graph shows textblob polarity score based on the rating we can see that average polarity score across different rating is increasing. Which shows that higher rating on average have higher polarity score.

# Findings : Textblob Polarity score for all brands

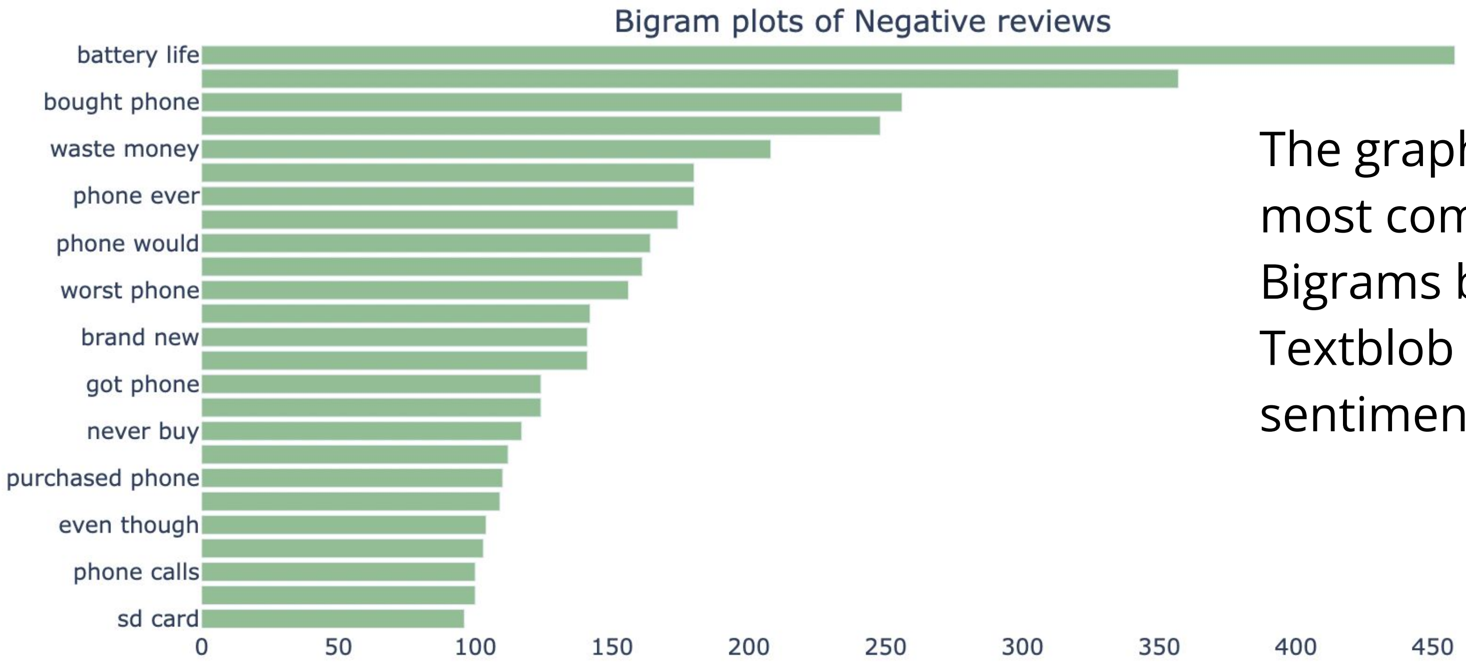


# Findings : Bigrams Positive Reviews




The graph shows the most common Bigrams based on the Textblob positive sentiment result

# Findings : Bigram Negative Review




The graph shows the most common Bigrams based on the Textblob Negative sentiment result.

# Limitations

As a result of above findings, this project provided answer to the research question by choosing the best sentiment analyzer based on the Naive Bayes Classifier. Moreover, other findings are that the battery life is the most common problem for the Negative sentiments.

In terms of limitation, the project disregarded emojis. Currently, emojis makes a large part of speech and comments or reviews. The sentiment analyzer provides best result when it is used for text. Thus this project did not consider reviews with emojis.

# Conclusions

In the project my main objective was to use different text analysis methods and evaluate the result based on the Naive Byes classifier, As a result I choose Textblob sentiment analyzer because it had the highest accuracy and had, better result in terms of precision, and f1 score.

We have seen that the most common problem of customers indicated from reviews is phones battery life and it is worth mentioning battery life is also source of the customer satisfaction. Thus, we can say phone battery is most significant part.

# Acknowledgements

One of the colleagues shared the following review:

“Text cleaning good! All in all very thorough project and well commented. Great job comparing two different text sentiment libraries. Good call using language detection to get rid of non-english review. Polished and insightful figures and visualisations. Could have looked at different classifiers and the parameterization of the vectoriser such as n-grams. Also your dataset is imbalanced you might have considered some ways to mitigate”

# References


- Halimedogan. (2021, November 13). *Sentiment Analysis (with NLP) for Amazon Reviews*. Kaggle. Retrieved July 24, 2022, from https://www.kaggle.com/code/halimedogan/sentiment-analysis-with-nlp-for-amazon-reviews
- Temoralkaisi. (2020, November 23). *Natural processing language*. Kaggle. Retrieved July 24, 2022, from https://www.kaggle.com/code/temoralkaisi/natural-processing-language/notebook#Apple
-