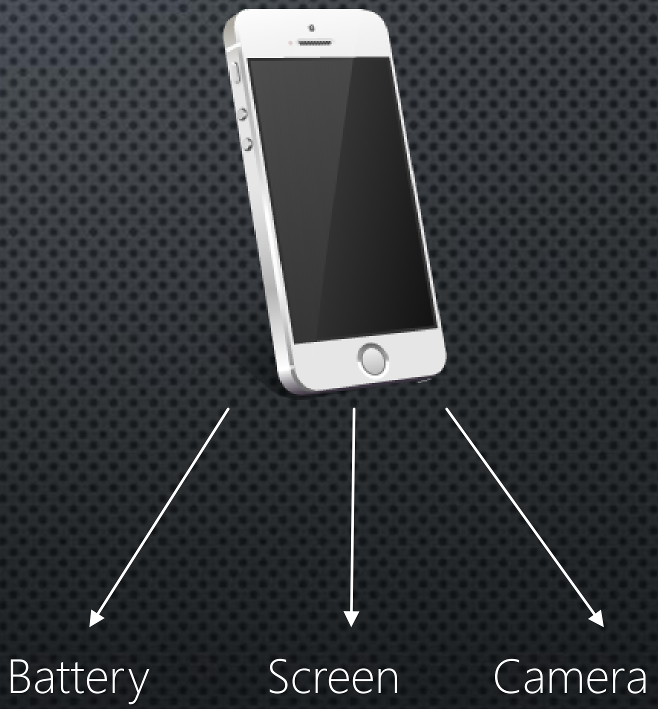
PRODUCT APPRECIATION ANALYSIS ON TWITTER VIA SENTIMENT ANALYSIS

ERALP BAYRAKTAR
ALİ UĞUR AKER
AYHUN TEKAT

OUR GOALS

- Creating a value for marketing by analyzİng real-time social media data
- Analyzing vast amount of data as efficient as possible
- Improving existing sentiment analysis approaches
- Customizing sentiment analysis for special domains

PLATFORM AND DOMAIN SELECTION



Battery

Screen

Camera

DIFFICULTIES

- Implementing a sentiment analysis from ground up
- Existing training sets for the analysis do not fit to our domain
- Getting real-time twitter data is expensive
- Need for an intelligent and efficient tweet extracting approach
- Spam tweets

Our solutions

- Switching to archİve data instead of using real-time data
- Tweet filtering
- Creating a sentiment analysis by using custom training data
- Using existing sentiment analysis library as a reference point

BIG DATA

- Archieve sizes:

  Chunk size per month: 50 – 60 GB (compressed)

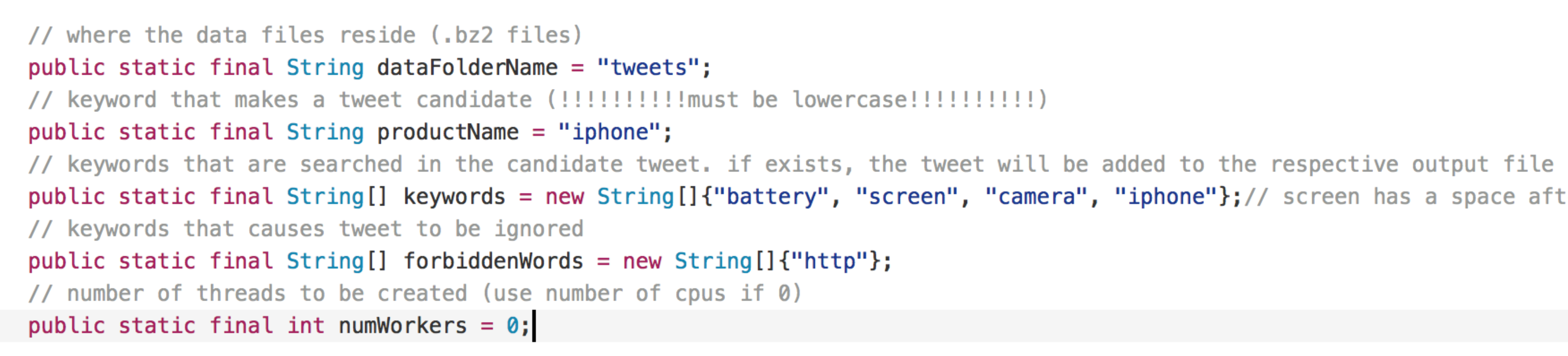
  150 GB(Uncompressed)

- Data from august - december 2014 is used
- Total data to process: 750 GB

750,000 x 💾

FILTERING TWEETS: TWEETCRAWLER

SETTINGS:



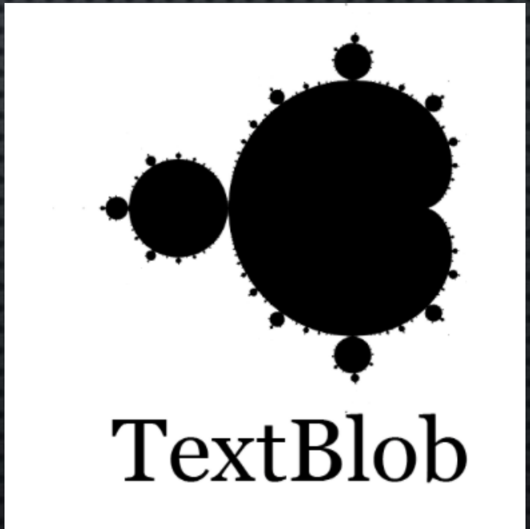
```java
// where the data files reside (.bz2 files)
public static final String dataFolderName = "tweets";
// keyword that makes a tweet candidate (!!!!!!!!!!must be lowercase!!!!!!!!!!)
public static final String productName = "iphone";
// keywords that are searched in the candidate tweet. if exists, the tweet will be added to the respective output file
public static final String[] keywords = new String[]{"battery", "screen", "camera", "iphone"};// screen has a space aft
// keywords that causes tweet to be ignored
public static final String[] forbiddenWords = new String[]{"http"};
// number of threads to be created (use number of cpus if 0)
public static final int numWorkers = 0;
```

FILTERING TWEETS: TWEETCRAWLER

- USER CAN CHOOSE NUMBER OF PROCESSING CORES
- PRODUCT NAME AND ITS DIMENSIONS CAN EASILY BE SET
- FORBIDDEN WORDS CAN BE DEFINED BY THE USER
- PROGRAM GOES THROUGH THE COMPRESSED FILES WITHOUT EXTRACTING THEM THEREFORE AVOIDS SUBSTANTIAL DISK RE QUIREMENTS
- 50% OF THE RUNNING TIME IS SPENT FOR ON THE FLY EXRACTING OF THE TWEETS

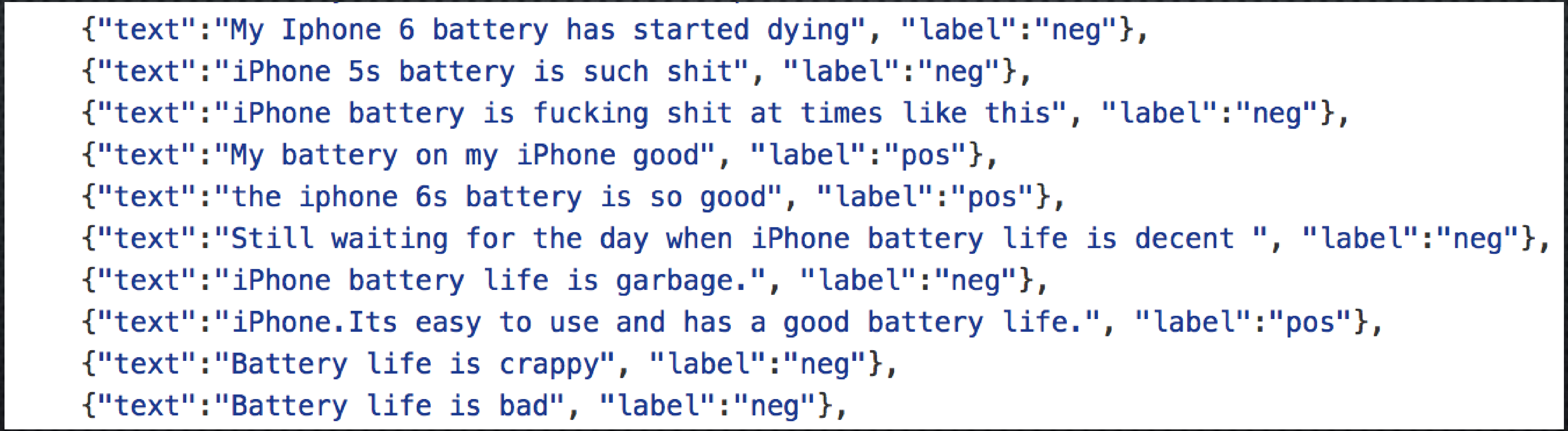
OUR SENTIMENT ANALYSIS APPROACH

- Naïve Bayes analysis + Emoticon analysis
- Textblob library as a reference point



NAÏVE BAYES ANALYSIS

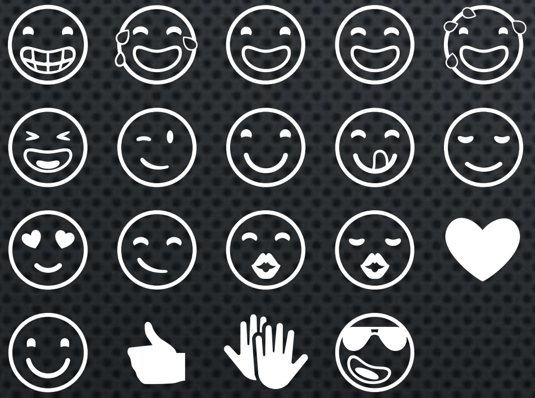
- CUSTOMIZED TRAINING SET
- CUSTOM PYTHON SCRIPT TO GENERATE TRAINING SET
- ALSO USED RECENT TWEETS IN OUR TRAINING SET



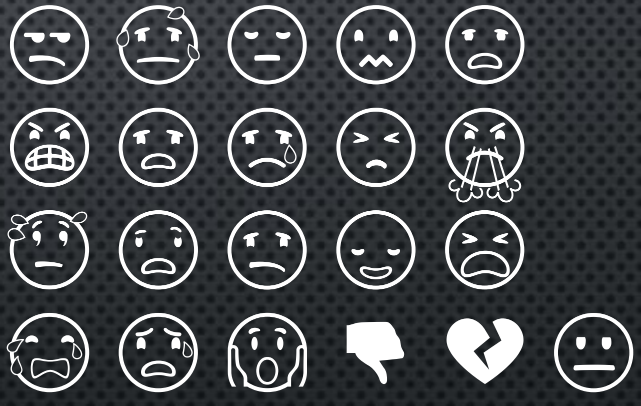
```
{"text":"My Iphone 6 battery has started dying", "label":"neg"},
{"text":"iPhone 5s battery is such shit", "label":"neg"},
{"text":"iPhone battery is fucking shit at times like this", "label":"neg"},
{"text":"My battery on my iPhone good", "label":"pos"},
{"text":"the iphone 6s battery is so good", "label":"pos"},
{"text":"Still waiting for the day when iPhone battery life is decent ", "label":"neg"},
{"text":"iPhone battery life is garbage.", "label":"neg"},
{"text":"iPhone.Its easy to use and has a good battery life.", "label":"pos"},
{"text":"Battery life is crappy", "label":"neg"},
{"text":"Battery life is bad", "label":"neg"},
```

EMOTICONS

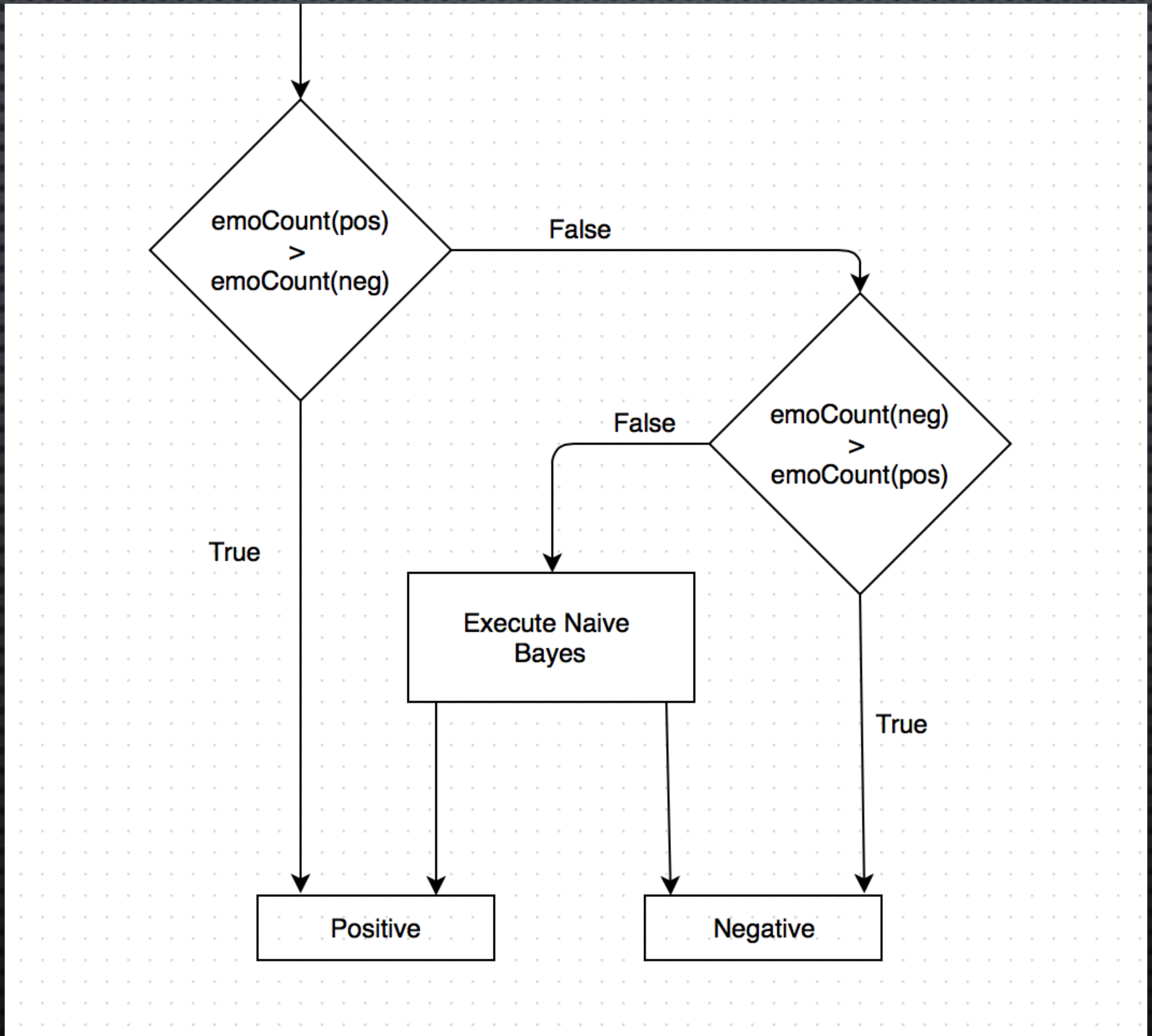
POSITIVE:



NEGATIVE:



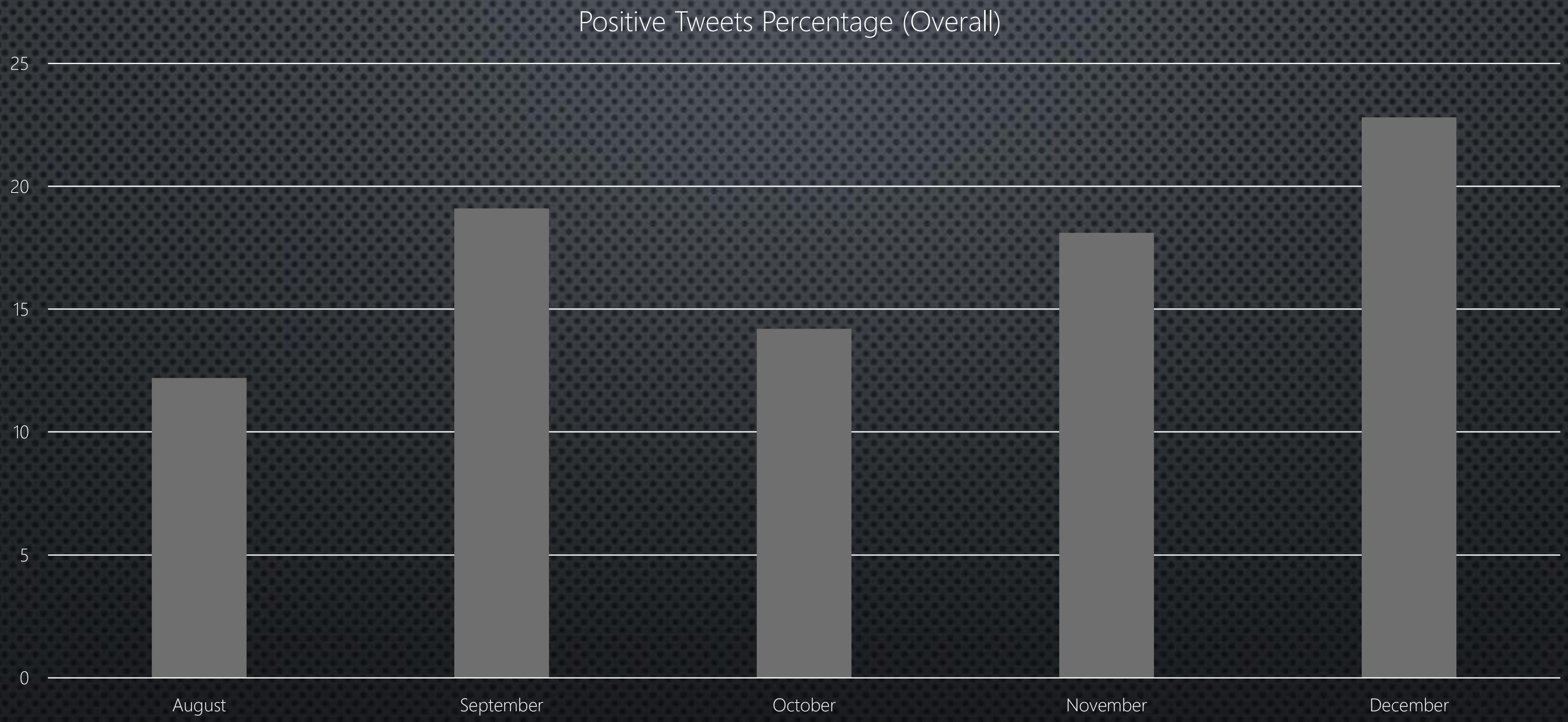
COMBINING: EMOTICONS + NAÏVE BAYES



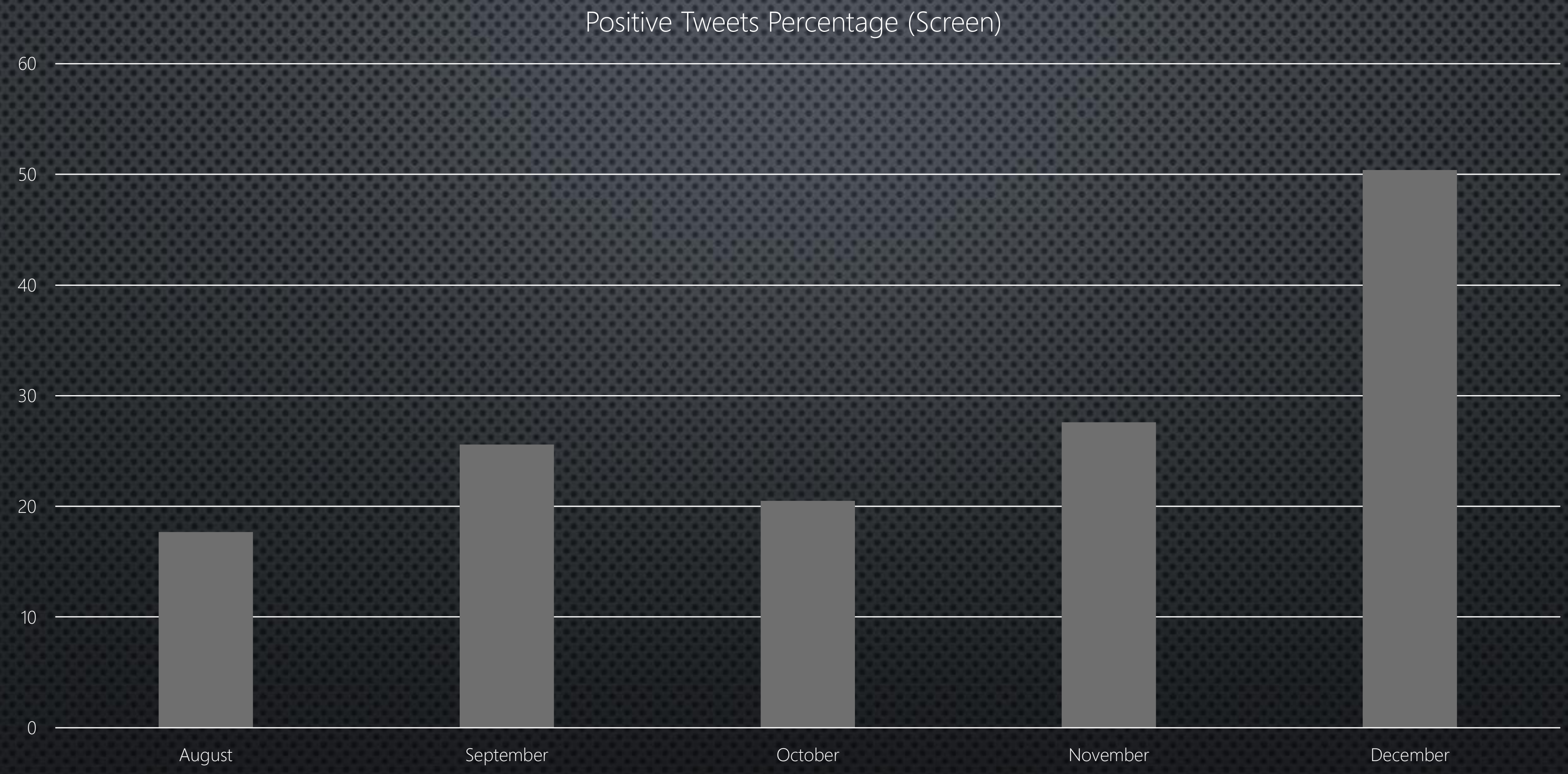
RESULTS

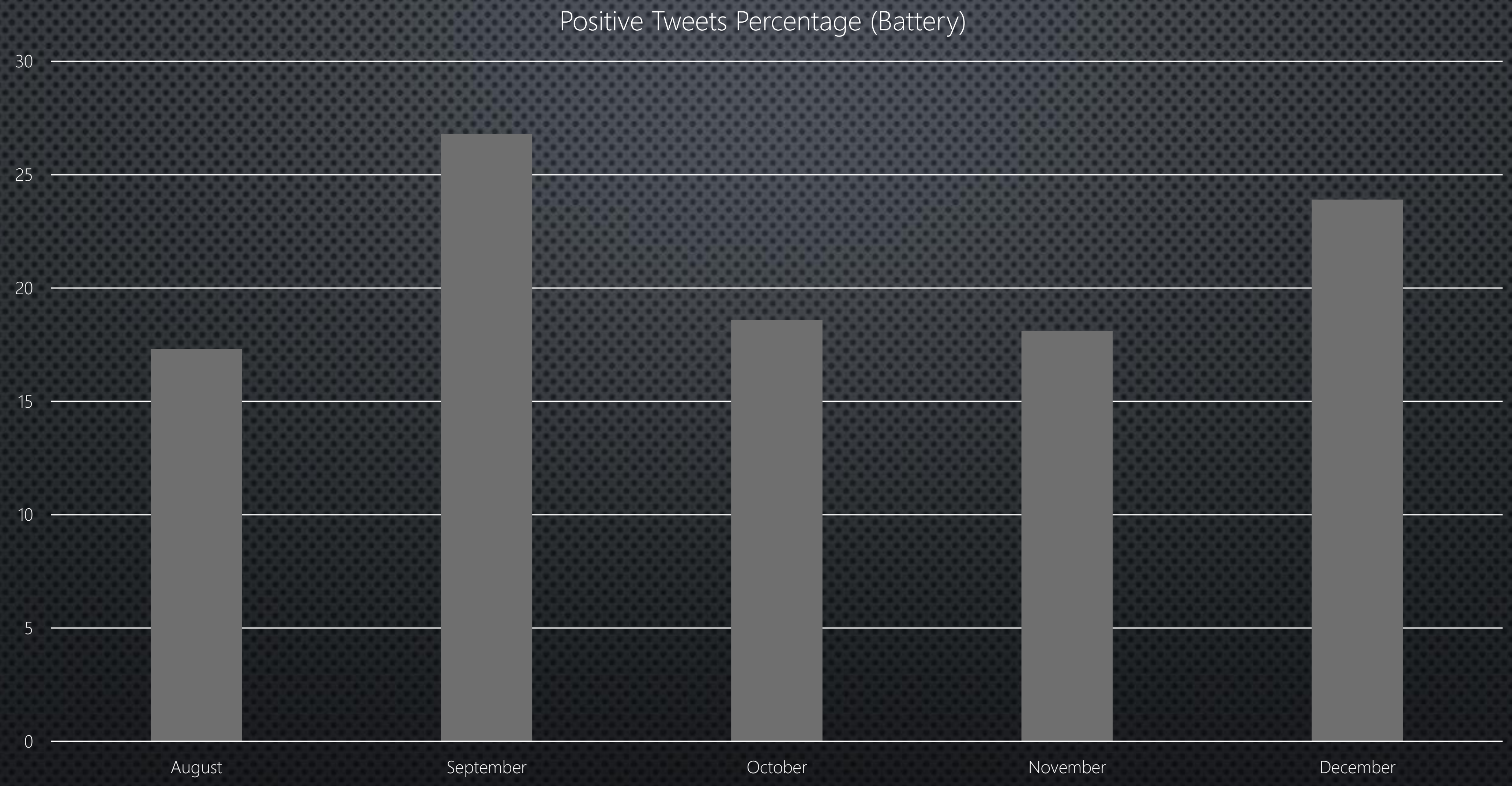
	TextBlob Positive	TextBlob Negative
Bayesian Positive	72.98%	15.70% (false positive)
Bayesian Negative	27.02% (false negative)	84.30%

	TextBlob Positive	TextBlob Negative
Emoticon Positive	78.00%	10.72% (false positive)
Emoticon Negative	22.00% (false negative)	89.28%

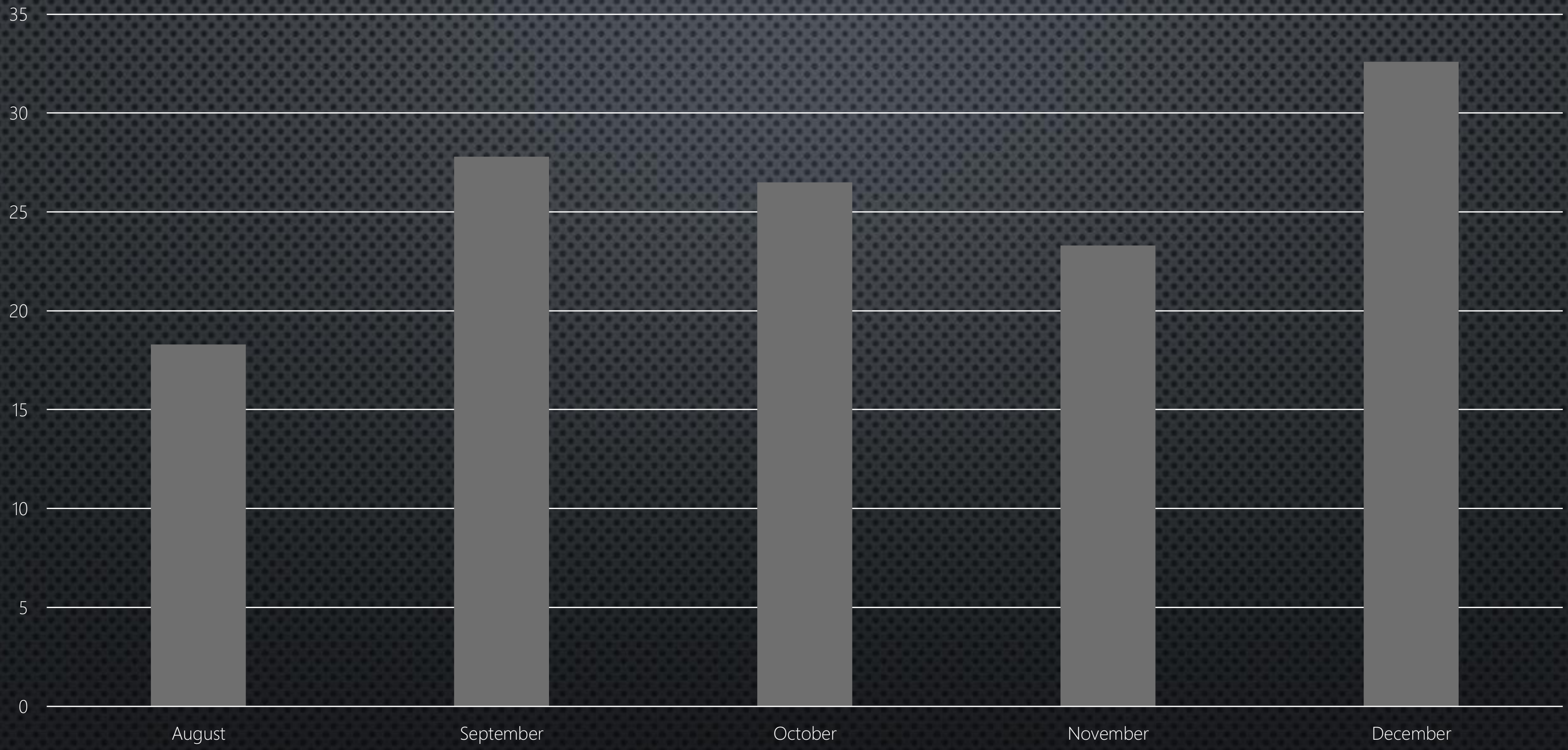


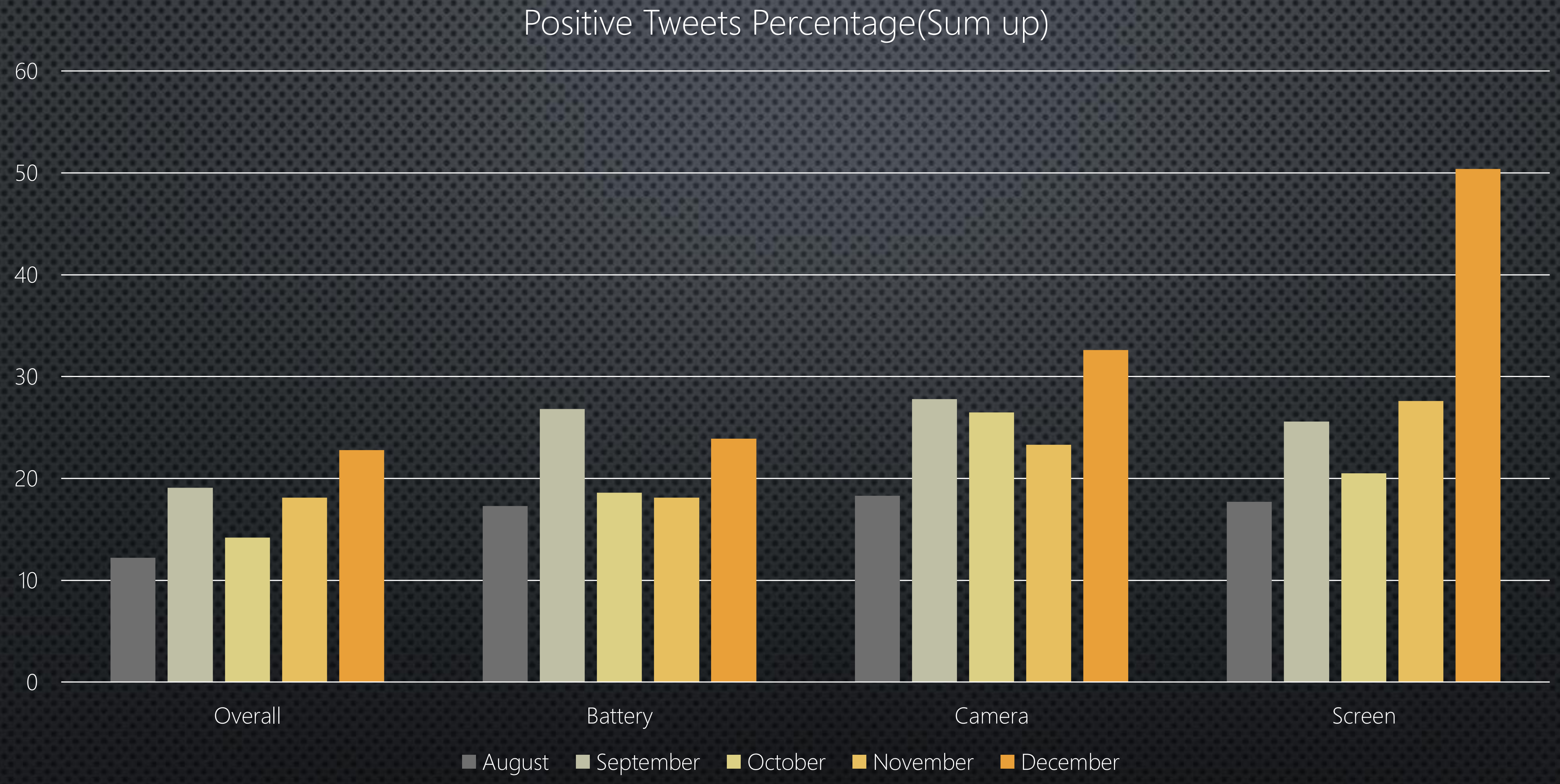
Positive Tweets Percentage (Overall)


Positive Tweets Percentage (Screen)
60
50
40
30
20
10
0
August
September
October
November
December

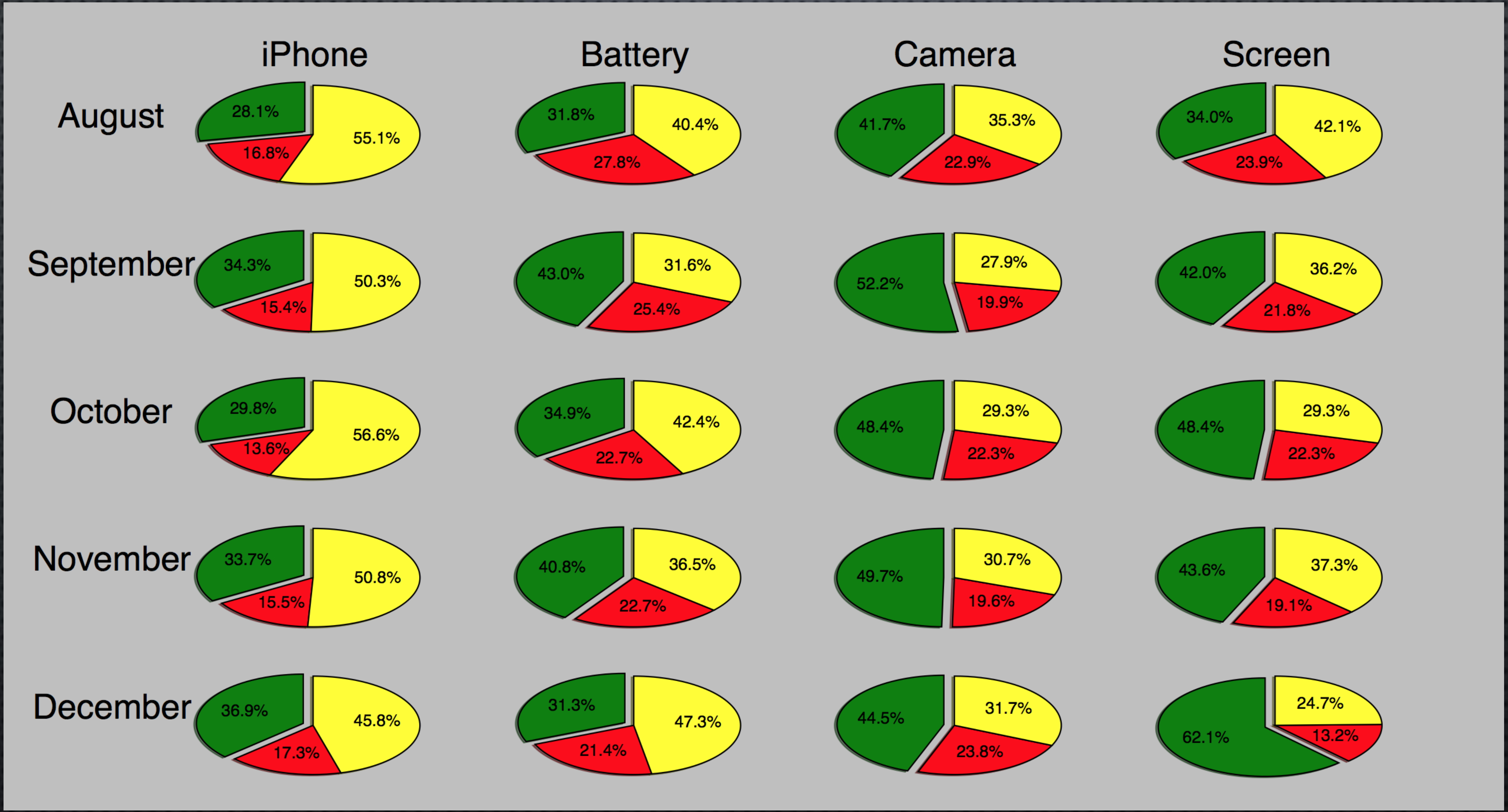

Positive Tweets Percentage (Battery)
30
25
20
15
10
5
0
August
September
October
November
December

Positive Tweets Percentage (Camera)

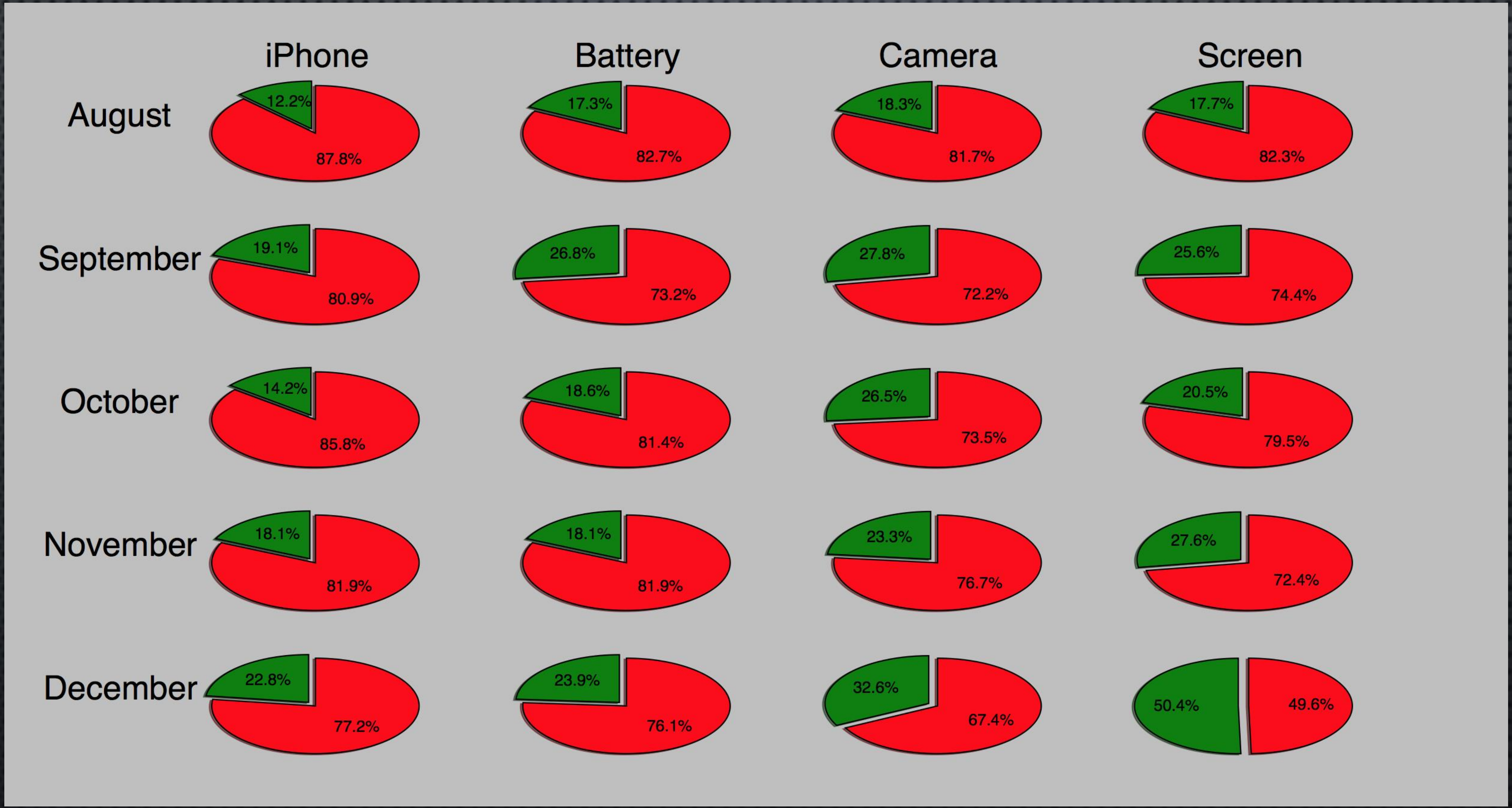



Positive Tweets Percentage(Sum up)
60
50
40
30
20
10
0
Overall
Battery
Camera
Screen
August
September
October
November
December

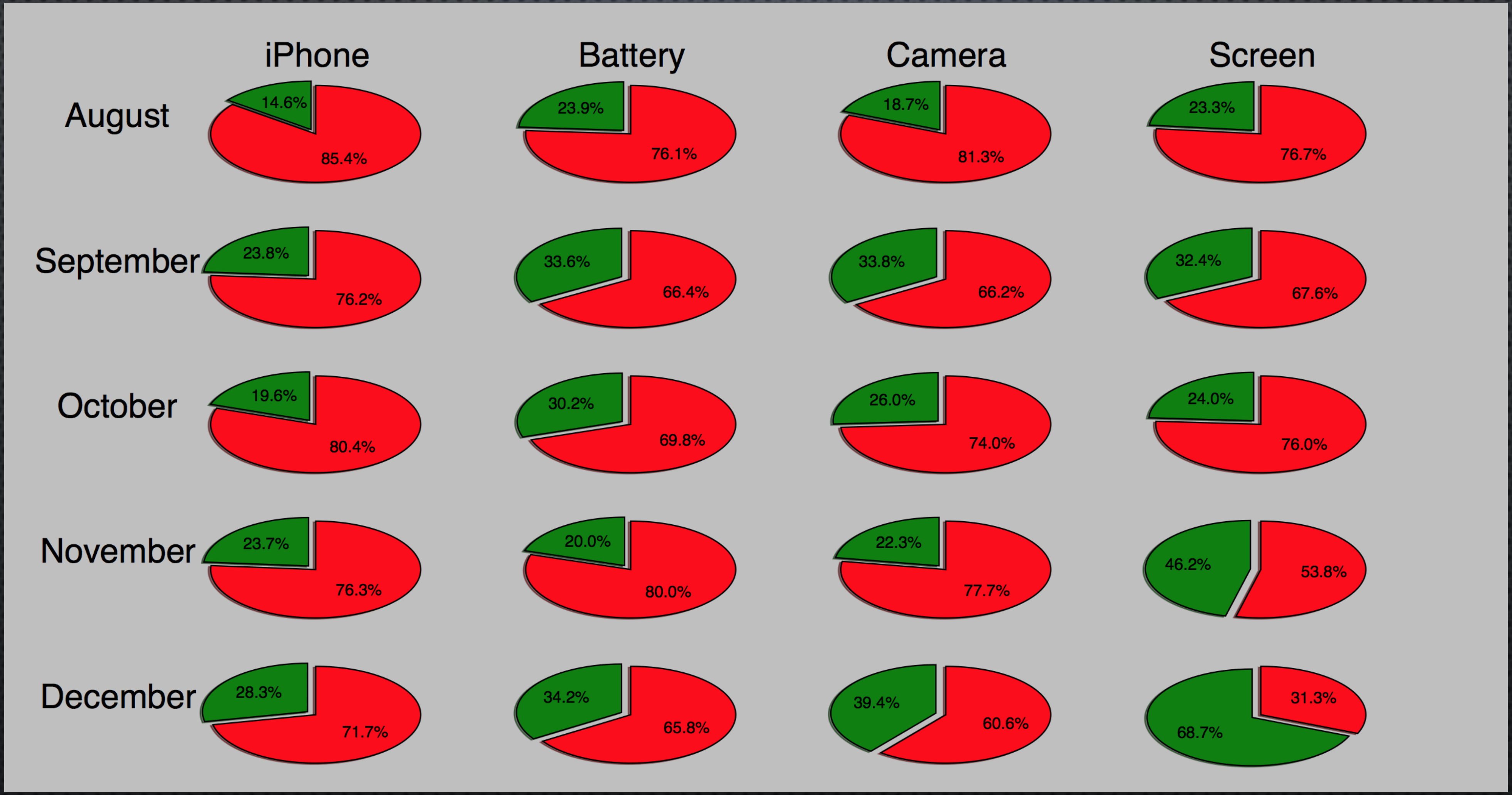
TEXTBLOB: PIG PICTURE



NAÏVE BAYES + EMOTICONS: PIG PICTURE



NAÏVE BAYES: PIG PICTURE



THANK YOU FOR LISTENING