

Bimodal Data Mining Report

Number of files analyzed: 4987

Number of text files analyzed: 2479

Number of audio files analyzed: 2508

Total file size analyzed: 24929144825 bytes (24.93 GB).

Total text file size analyzed: 28305493 bytes (0.03 GB).

Total audio file size analyzed: 24900839332 bytes (24.9 GB).

Total data word count: 28591665

Most common words:

like (20349)

peopl (19963)

go (18463)

think (17747)

know (16228)

thing (14974)

year (13406)

time (12907)

look (11919)

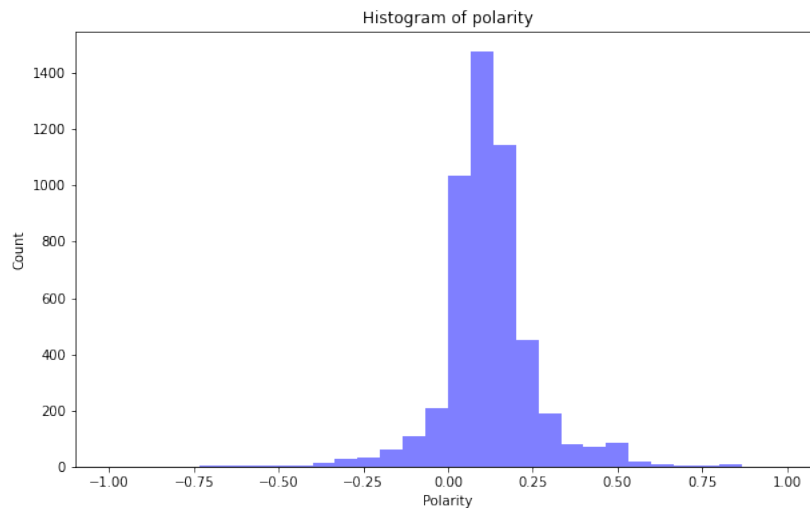
want (11410)



Average dataset polarity: 0.115

Polarity is a representation of whether the emotion of the data is overall generally positive or negative, with -1 being a completely negative score, and positive 1 being completely positive.

Number of files per polarity subset:



Topic Breakdown

These wordsets represent the topics most representative of the entire dataset, based on their frequency of appearance. The words in each topic are shown with their relative weight of importance.

Topic 1:

0.013*"go" + 0.010*"year" + 0.009*"like" + 0.009*"thing" + 0.008*"think" + 0.007*"look" + 0.007*"peopl" + 0.007*"know" + 0.007*"come" + 0.006*"want"

Topic 2:

0.012*"peopl" + 0.008*"year" + 0.008*"think" + 0.008*"like" + 0.007*"world" + 0.007*"go" + 0.007*"thing" + 0.006*"time" + 0.006*"know" + 0.006*"countri"

Topic 3:

0.015*"like" + 0.011*"laughter" + 0.010*"know" + 0.010*"thing" + 0.009*"go" + 0.009*"time" + 0.008*"think" + 0.007*"look" + 0.006*"want" + 0.006*"work"

Topic 4:

0.008*"like" + 0.008*"peopl" + 0.007*"look" + 0.007*"come" + 0.006*"say" + 0.006*"world" + 0.006*"know" + 0.006*"go" + 0.006*"year" + 0.006*"build"

Topic 5:

0.012*"like" + 0.011*"go" + 0.009*"thing" + 0.009*"know" + 0.009*"think" + 0.009*"actual" + 0.008*"peopl" + 0.007*"work" + 0.007*"time" + 0.006*"look"

Topic 6:

0.014*"peopl" + 0.012*"think" + 0.010*"like" + 0.010*"go" + 0.009*"know" + 0.009*"thing" + 0.007*"say" + 0.007*"time" + 0.007*"want" + 0.007*"year"

Average Emotional Distribution:

