

Classifying Google Home & Alexa Subreddits

By: Joshua Mizraji



Setting The Scene

Google

amazon

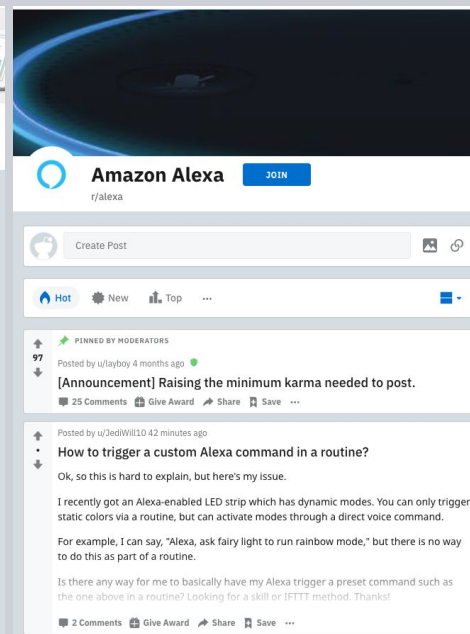
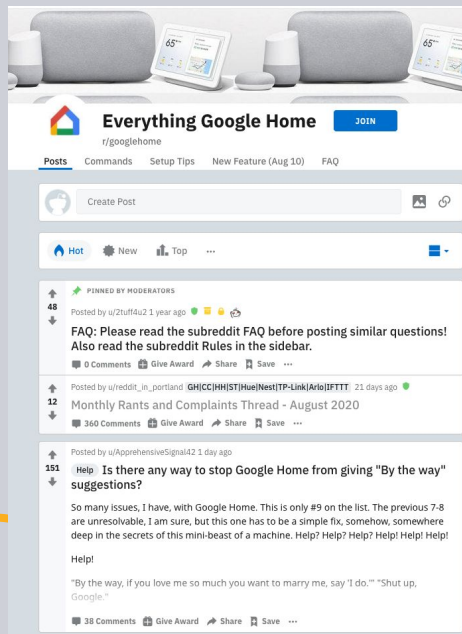


Let's imagine that I'm an outside consultant tasked with analyzing
Reddit data in order to better understand customer feedback
regarding Alexa and Google Home devices.

[Image Source](#)

Problem

Reddit data is text heavy, which makes reading through each post too time consuming of a task for a human.





[Image Source](#)

Objective

My objective is to use **NLP** to compare subreddit content, which will be used to **train a classifier** on which subreddit a given post came from.

By assessing the data, we can understand what users are saying about each company's devices and **provide better business recommendations.**

Data Exploration

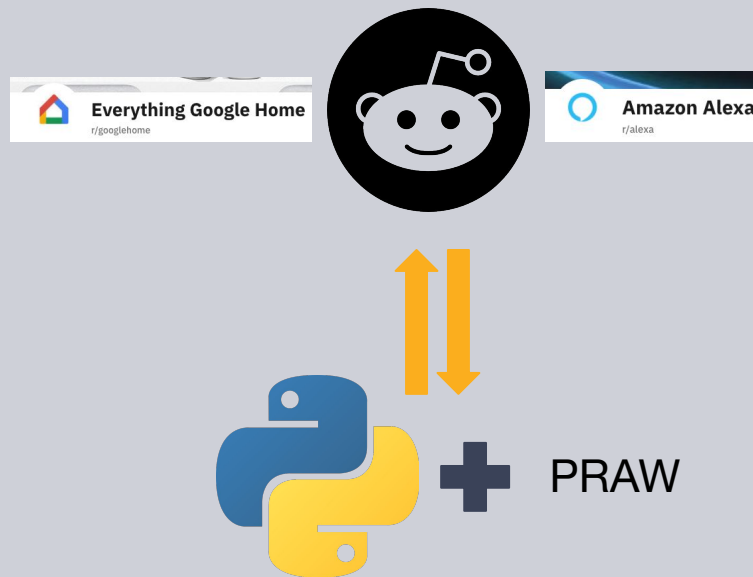
The Data

Alexa Subreddit

- 39.9k Members
- Created on Feb 19, 2009
- 998 Posts
- 4482 Comments

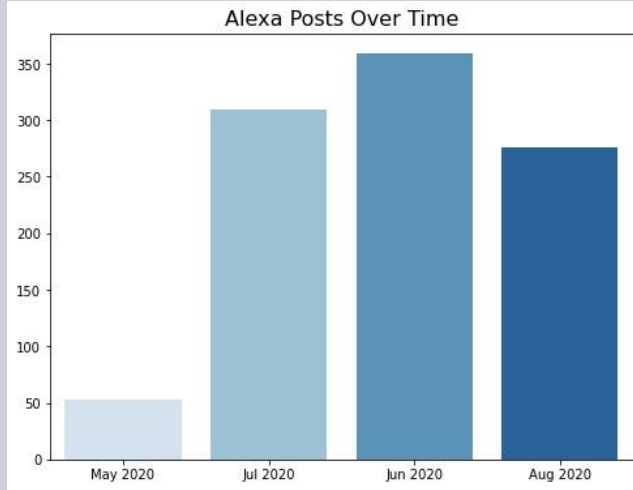
Google Home Subreddit

- 186K Members
- Created on May 18, 2016
- 968 Posts
- 6233 Comments

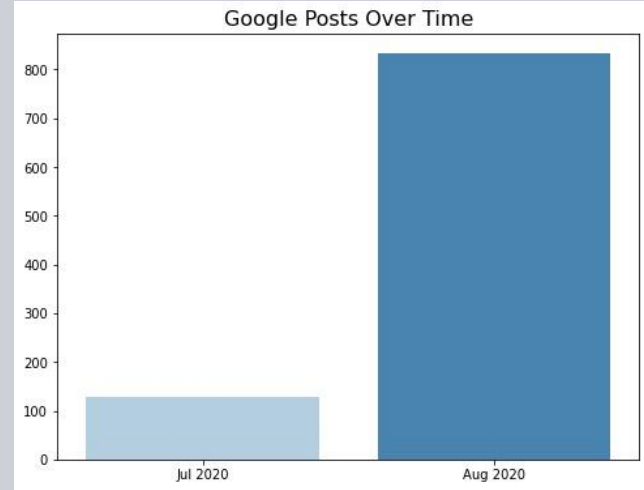


Scraped the Reddit API using [PRAW](#)

Posts Over Time

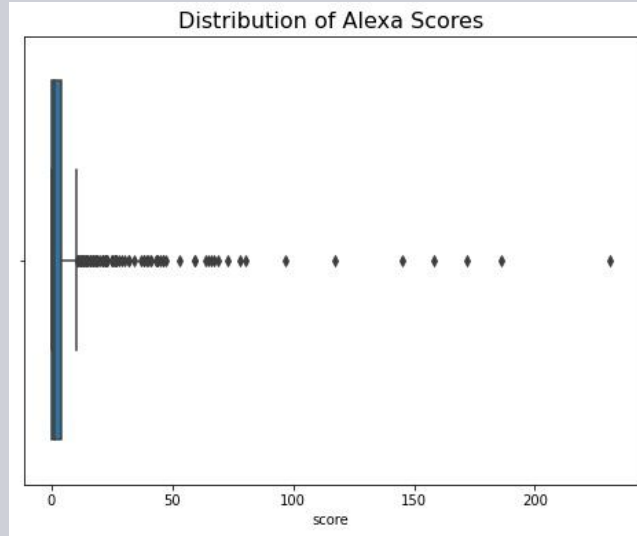


Alexa

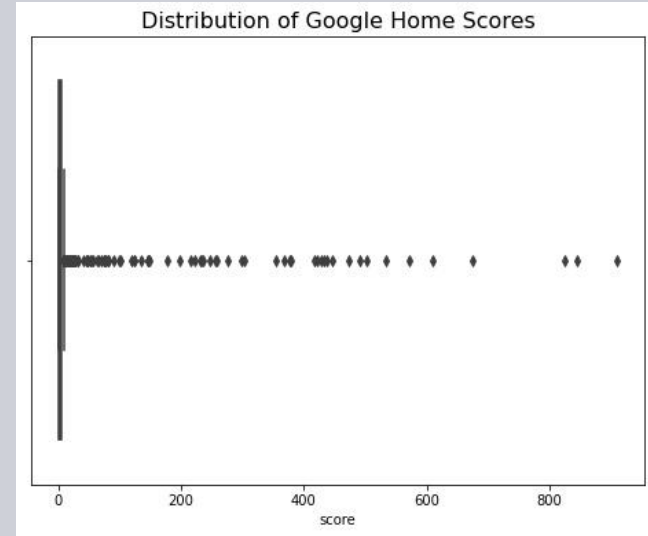


Google Home

Distribution of Scores

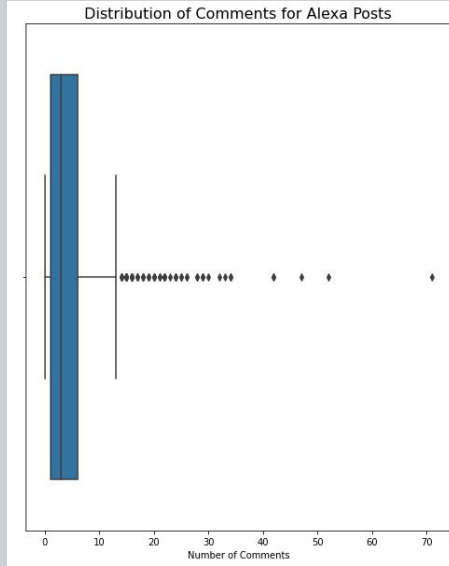


Alexa

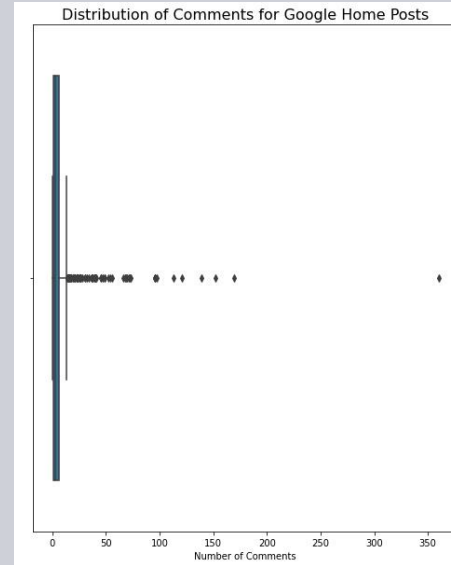


Google Home

Distribution of Comments



Alexa



Google Home



Alexa Posts With The Highest Score

“I created a skill that will play “Piano Man” by Billy Joel so it will play the line “it’s 9 o’clock on a Saturday” at exactly 9:00 on a Saturday.”

“the alexa app is crap. it's slow, laggy, fully of bugs, and just a pain in the arse to use.”

“My kids have worked out how to set reminders on their siblings Echo from their own”

“I am currently working on an Alexa skill that lets users play YouTube videos via audio on Alexa devices without a screen/ plays the video on those devices with a screen. Essentially this is a skill that lets users listen to music/videos for free without a subscription! “

“Anyone Else Get Really Annoyed by Alexa’s “By the way...”

“A new feature apparently!”

“Alexa getting less intelligent all the time.”





Google Home Posts With The Highest Score

“Dark mode has arrived, at long last!”

“I got rick rolled by Assistant “

“I really should be more specific.”

“Cat just threw up and I was telling them to stop doing it and Google laid this on me”

“Today, my kid accidentally dropped and broke glassware and then I received this message. I was happy it recognized. Should I be concerned?? I read somewhere that Google Assistant would not process until it's prompted with " Ok Google"

“Do yourself a favor and ask google “what sound does a camel make?”

“My partner was drunk and home alone last night and...”

“Google just made my dumb smoke detectors smart”



What does this tell us?

Funny images and videos get the most upvotes.
Feature requests and collective user frustrations
also get upvoted.



Alexa Posts With The Most Comments

“Alexa has my smart light colors all wrong? “

“Can i get her to stop announcing that she's now connected to bluetooth?”

“Favorite Automation or Routine?”

“Amazon Basics smart microwave - my first impressions”

“Got my Alexa glasses today!“

“Got my Echo Auto today in the UK. Can't wait to get to use it.”





Google Home Posts With The Most Comments

“Monthly Rants and Complaints Thread - August 2020”

“Google Play Music will be discontinued starting September. New Zealand and South Africa you go first.”

“PSA: Per Google, playlists for YouTube Music is not yet supported. I sure hope this happens before they sunset GPM, which is supposed to happen sometime this year.”

“Can anybody tell me what is going on in this video? I have several clips of this taken last night (8/18) it looks like snow but it's summer lol”

“Casual Ad-Free Music on GH Devices in The Future”

“Google Makes Big Bet On ADT”

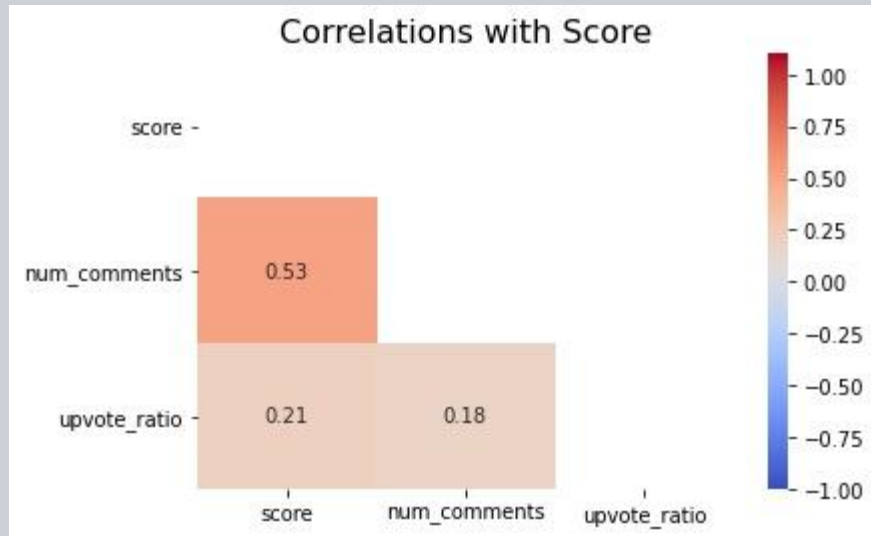


What does this tell us?

Posts with the most comments are similar to posts with the most upvotes, in that they provide new product announcements, favorite functionality, and frustrations with the product.

Correlations

There seems to be a slight positive correlation between the number of comments and score.



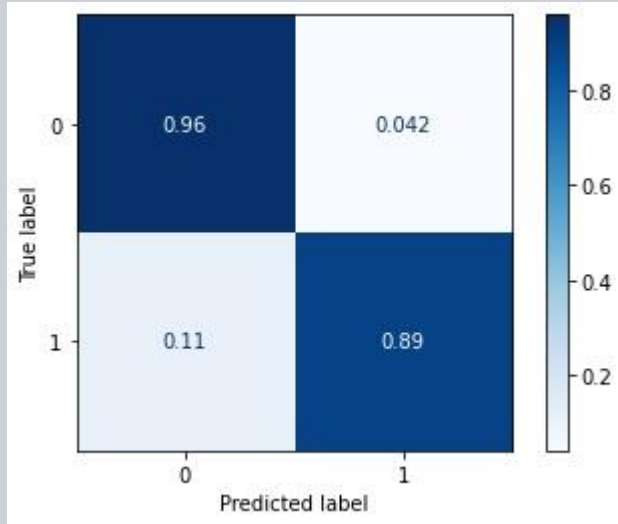
Modeling



50%

The Baseline score to beat

Logistic Regression



Confusion Matrix

99.6%

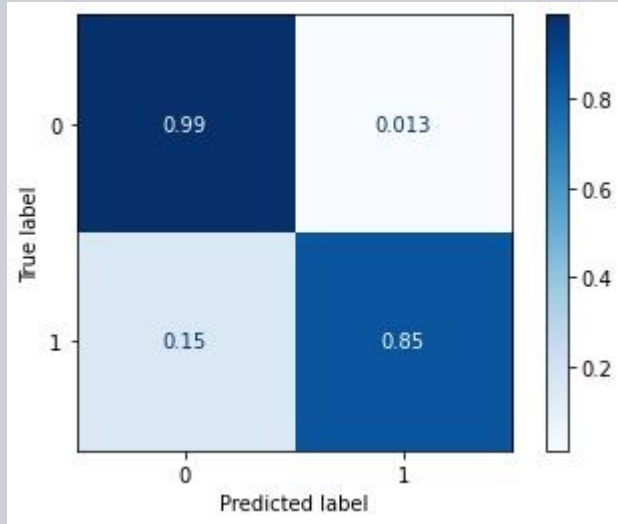
Training Accuracy

92.4%

Testing Accuracy

Used CountVectorizer with english stop words and Gridsearch to optimize my hyperparameters

Random Forest



Confusion Matrix

94.8%

Training Accuracy

91.6%

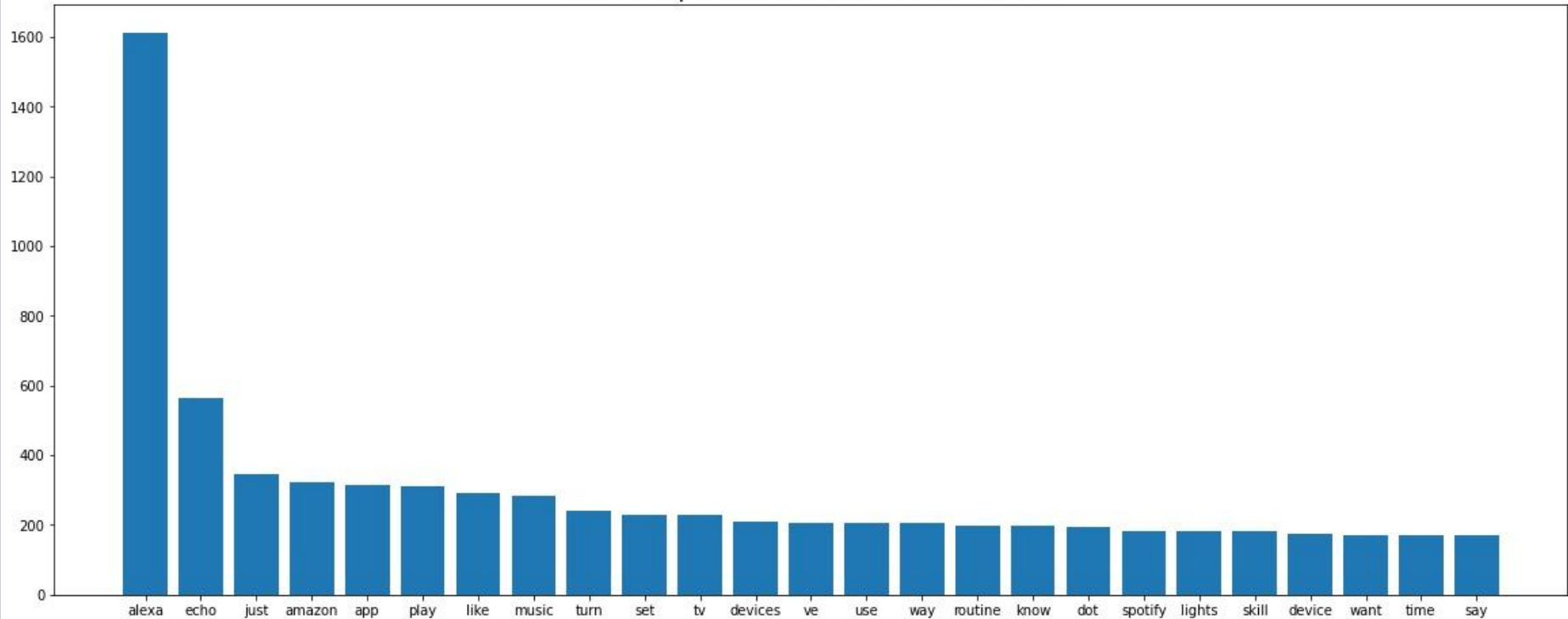
Testing Accuracy

Used CountVectorizer with english stop words and Gridsearch to optimize my hyperparameters

Our models are both performing above 90% accuracy, let's investigate why?

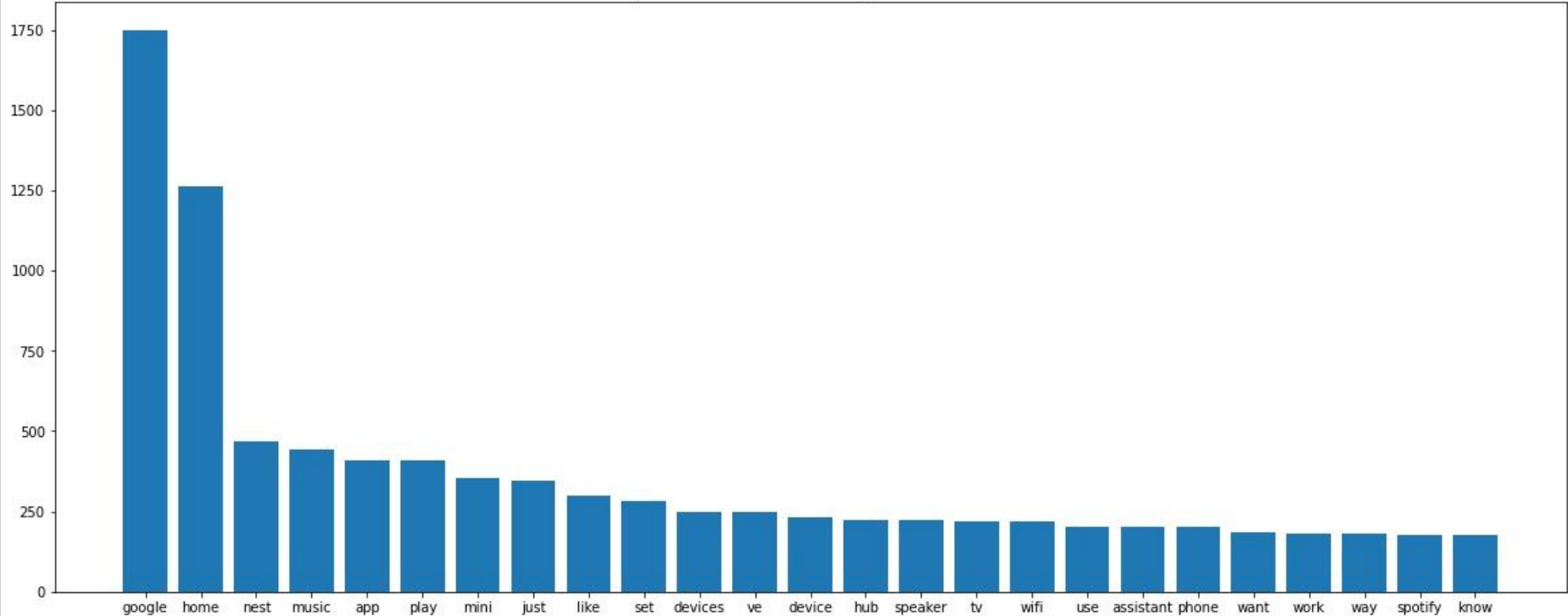
Alexa Most Frequent Words

Most Frequent Words in Alexa Subreddit



Google Home Most Frequent Words

Most Frequent Words in Google Home Subreddit



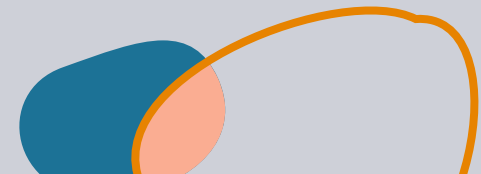
Since the words “alexa” and “google” show up so frequently in the posts, let's see if we can make it more challenging to predict the class by removing some of the most unique popular words.



New Stop Words

We removed the followed words from the analysis:


'alexa', 'echo', 'dot', 'google', 'home', 'nest',
'assistant'





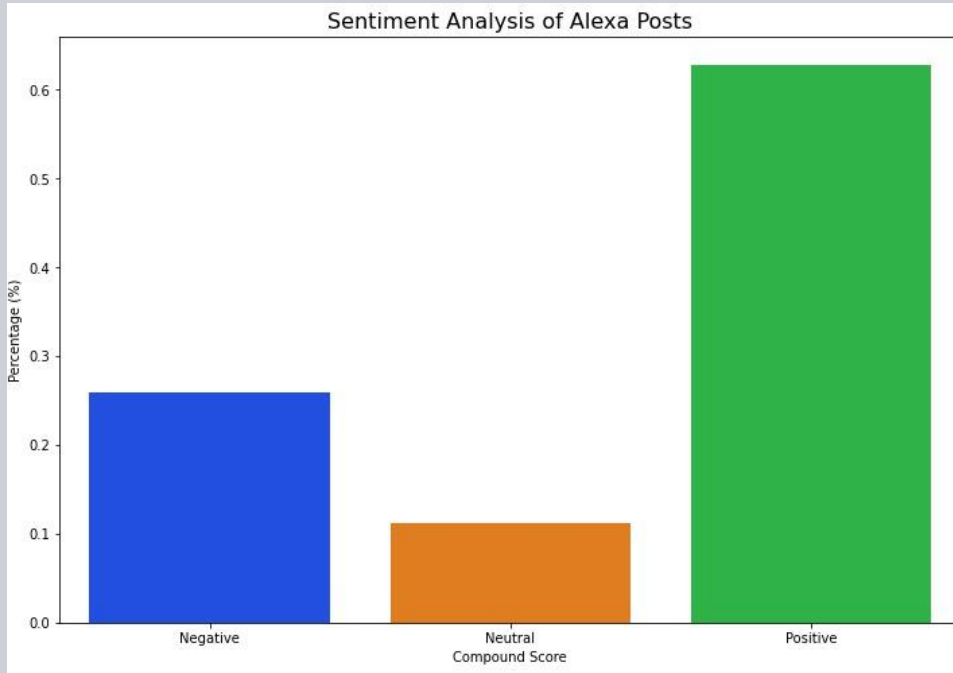
Updated Models

	Training Accuracy	Testing Accuracy
Logistic Regression	99.3%	79.6%
Random Forest	71.9%	69.6%
Multinomial Naive Bayes	92.4%	73.5%
Random Forest (Bagging + AdaBoost + Voting Classifiers)	87.7%	75.9%



Sentiment Analysis

Alexa Sentiment Analysis of Post Title & Body



62.9%

Positive

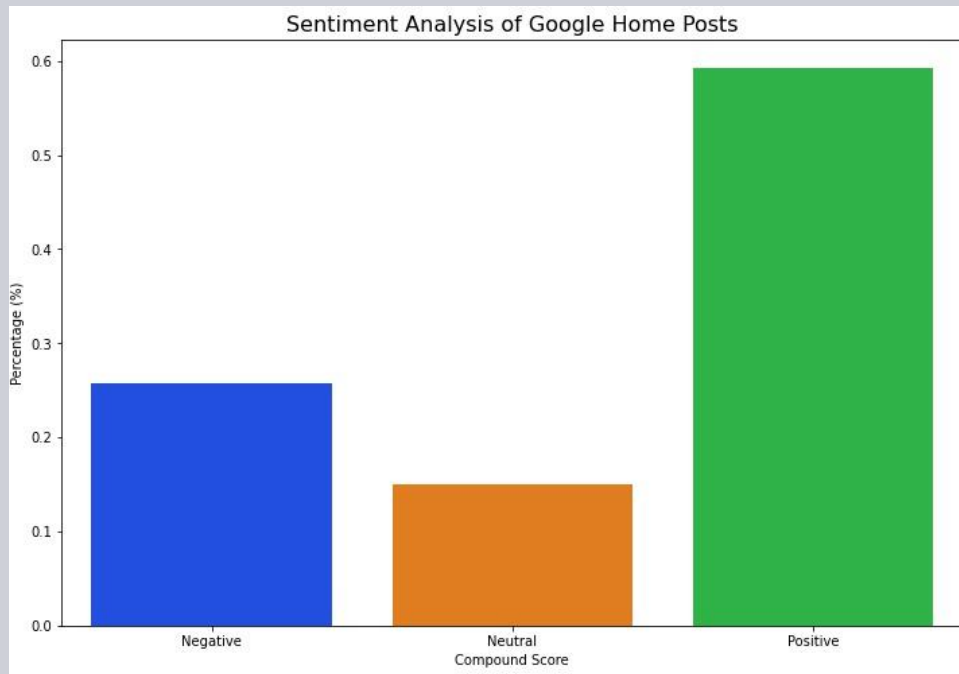
25.9%

Negative

11.2%

Neutral

Google Home Sentiment Analysis of Post Title & Body



59.3%

Positive

25.8%

Negative

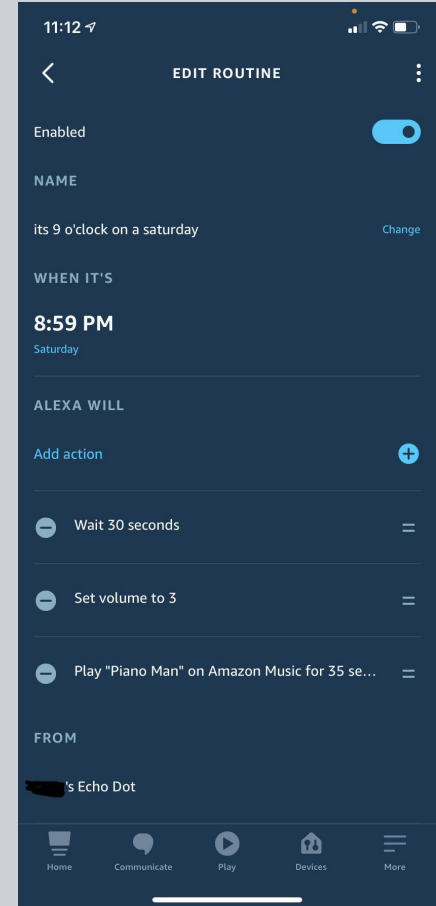
14.9%

Neutral

Business Recommendations

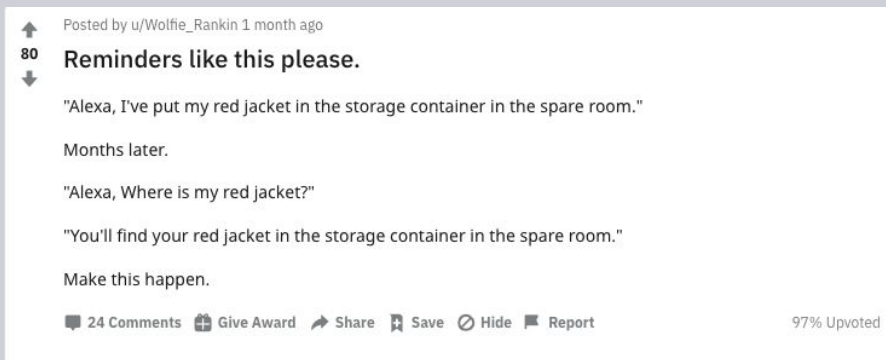
Humor Gets Engagement

Utilize funny posts in your social media and marketing. Showcase the innovative and humorous things users are creating with the devices.



Consider Feature Requests

Creative ideas can come from anywhere. Explore feature requests to find creative ideas for new features and skills.



Listen To User Frustrations

Users will switch to another product if their needs are not met. Consider pitfalls, engage with these users, and improve the overall experience.



NefariousnessSolid87 14 points · 19 days ago



The Google ecosystem is becoming less attractive by the day. Forcing us into YTM was annoying but it was something I figured I could adjust to but the fact that routines have been broken for several days now without even the slightest acknowledgement from Google is a bridge too far for me. All I want is for a specific playlist to play at a specific time on a specific speaker. This is apparently not possible with Google anymore. I'm really tired of being Google's test subject. I'm supposed to be your customer, not your guinea pig. They push an untested update that breaks routines, then when it's brought to their attention, they can't be bothered to roll it back and their new music service is barely functional with their own Home products.

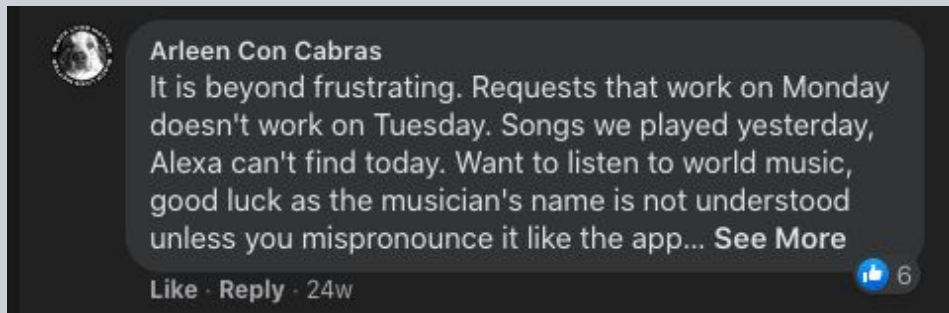
I've got an Amazon Echo on the way. If it's suits my needs then I'm done with this. All my nests and minis will be stowed in a closet and I guess I'll have to subscribe to Apple Music.



Reply Give Award Share Report Save

Utilize NLP With Additional Outlets

Apply similar analysis to App store reviews, product reviews, and social media comments.





Thank You!

CREDITS: This presentation template was created by **Slidesgo**, including icons by **Flaticon** and infographics & images by **Freepik**.

Please keep this slide for attribution.

