Karam Said

FMBAN2 2021 – 2022

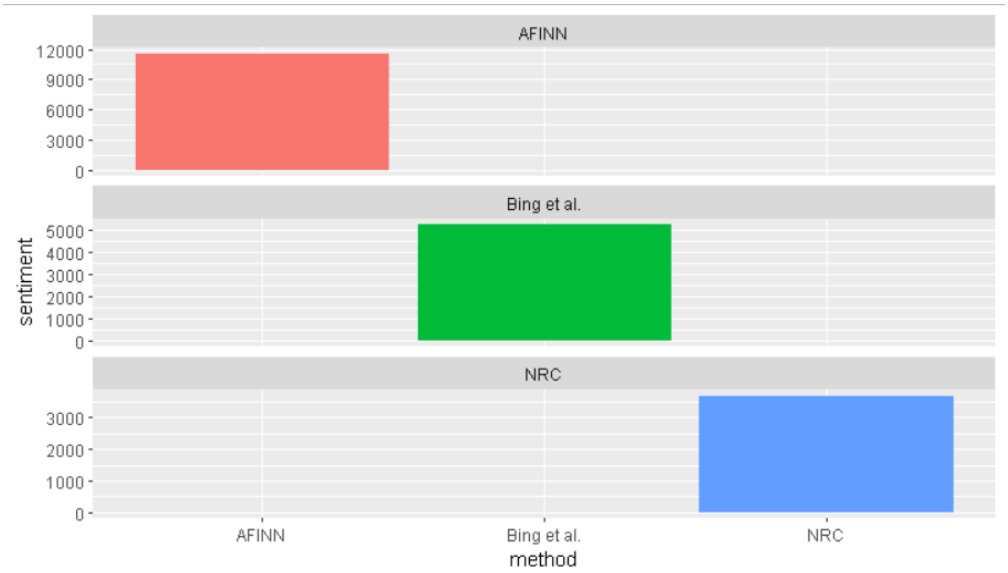
Text Analytics and Natural Language Processing (NLP)

12/5/2021

Business Insight Report

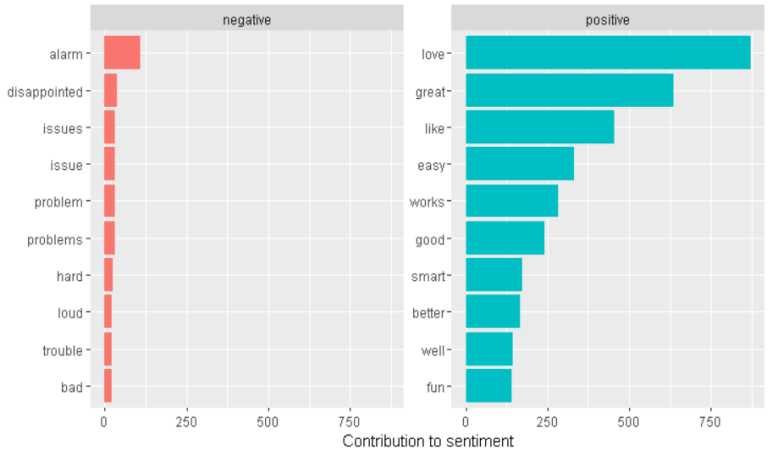
This dataset is a variety of Amazon Alexa products such as Echo and Echo Dot with 3000 Amazon Customer Ratings , Star Ratings and Rating Dates. In order to analyze the customer reviews in our report, we will opt three different frameworks which are the Sentiment Anaysis framework, N-grams Analysis framework and the TF-IDF Analysis framework.

Sentiment Analysis



After running the sentiment analysis, we can see strong positive sentiments with all of our three methods (AFINN, BING and NRC). That means that most of the reviews are positive reviews. But to be more precise and have more insights on the two different products, we found the most common positive and negative words for each.

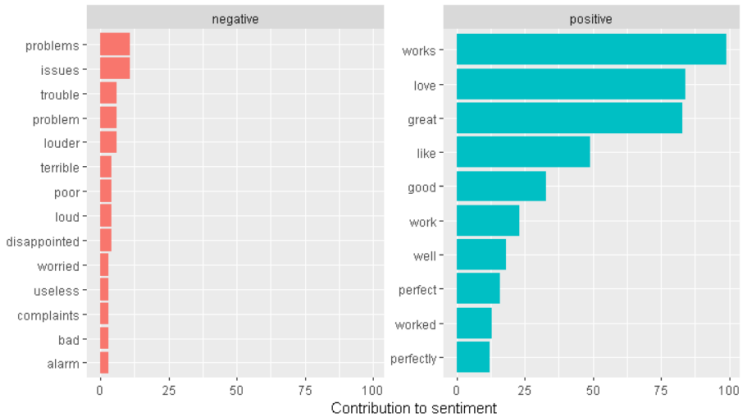
Most common positive and negative words for Echo



(In our context, “alarm” isn’t considered as a negative word so we’ll just ignore it)

For the Amazon Echo, we can see that we clearly have more positive than negative words. The most common words in the reviews are “love” and “great” which shows that in general everything about the product is good. We still have some negatives and diappointments probalby related to small issues.

Most common positive and negative words for Echo Dot

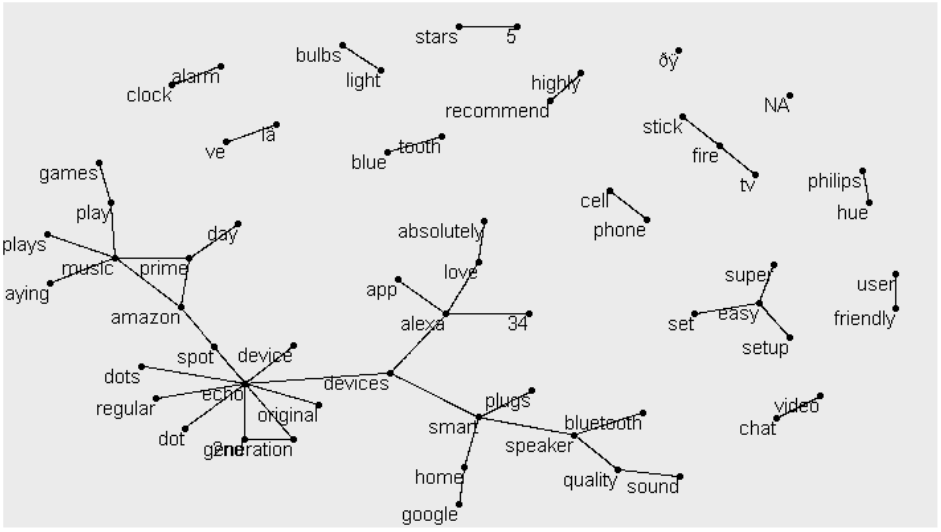


Additionally, for the Amazon Echo Dot, we can also see that we mainly have positive words that shows that this product is great in general. But on the other side, we can tell that the two most common negative words are “problems” and “issues”. When looking more into the words, we also find the word “loud” or “louder”, that means that most probably the main issue is with the volume and maybe it could go a bit more higher. The word “poor’ shows that som of the reviews says that audio quality is poor and could be better.

In conclusion, we can tell that two products are doing great but we have some negative reviews for the Echo Dot that explains that the customers were expecting the volume to be higher.

N-Grams Analysis

In this analysis, we paired the words into bigrams so we can get more buisness insight.



From this bigram network above, we can see the bigram “highly recommend” which actually shows recommendation in some of the reviews and the bigram “5 stars” that gives it a good rating. Moreover, we can also tell the product is userfriendly, it’s super easy to use and it is easy to set It up.”love absolutely” tell us that the Amazon Alexa Echo products are really loved by the customers and shows how it actually integrated in their life routine with its reminders and alarms. In addition, “play games” indicates that the game feature that the smart speaker has was a great addition. Additionally, the bigram “prime day” shows that the product could have been bought on a discount on prime day. That shows the importance of prime day and how much Amazon sells items on prime day. Discounts on their own products gives them an advantage to sell more. Also, we can see the bigram amazon prime music that shows its importance as a additional subscription with the product. People who are not subscribed with amazon music before buying an Echo product, will be more likely to subscribe after buying the product because the premium version will give you access to all of the songs and not just limited to certain songs. This shows the example of complemetary products, which means buying one of them could actually result in buying the other one.

TF-IDF Analysis

In the TF-IDF analysis, we decided to create quadrograms to get more business insights. We can see in the screenshot down below in row 1, the highest tf-idf score which is about 2.77 shows the quadrogram “prodcut wow family fun”. This can give tell that it is a really good family friendly product. The easy access of the speaker with its voice assistant feature is shown in row 12 with the quadrogram “plays music easily accessible”. This shows how this smart speaker differentiates itself from any other normal speaker on the market. In row 4, we can see ”home security system” which shows the importance of Alexa Guard, “which can detect if someone is breaking into your house, and Alexa Drop-In, which lets you listen in to see what the detected noise could be”. On the other hand in row 2, we can see the quadrogram “echo stops playing music” with a tf-idf score of 1.38. This shows that there is some kind of issue with the product where the music stops on its own.

	variation	quadrogram	n	tf	idf	tf_idf
1	Walnut Finish	product wow family fun	2	1.00000000	2.772589	2.77258872
2	Oak Finish	echo stops playing music	2	0.50000000	2.772589	1.38629436
3	Oak Finish	playing music setting timers	2	0.50000000	2.772589	1.38629436
4	Heather Gray Fabric	abode home security system	2	0.09090909	2.772589	0.25205352
5	Heather Gray Fabric	adding smart home devices	2	0.09090909	2.772589	0.25205352
6	Heather Gray Fabric	alexa compatible power strips	2	0.09090909	2.772589	0.25205352
7	Heather Gray Fabric	alexa play games play	2	0.09090909	2.772589	0.25205352
8	Heather Gray Fabric	cool product speaker sounds	2	0.09090909	2.772589	0.25205352
9	Heather Gray Fabric	games play radio stations	2	0.09090909	2.772589	0.25205352
10	Heather Gray Fabric	news weather information itâ	2	0.09090909	2.772589	0.25205352
11	Heather Gray Fabric	play games play radio	2	0.09090909	2.772589	0.25205352
12	Heather Gray Fabric	play music easily accessible	2	0.09090909	2.772589	0.25205352
13	Heather Gray Fabric	play radio stations play	2	0.09090909	2.772589	0.25205352
14	Heather Gray Fabric	purchased alexa compatible power	2	0.09090909	2.772589	0.25205352

R CODE

```
> ####################################################################
> ### Sentiment analysis for Alexa ######
> ####################################################################
>
> library(tidytext)
> library(dplyr)
> library(stringr)
> library(tidyr)
> library(tidytuesdayR)
>
> colnames(amazon_alexa)[4] <- "text"
>
> #creatng the new variable for the product because we only have the color variation
>
> amazon_alexa$product <- c()
> for (i in 1:nrow(amazon_alexa)) {
+   if (amazon_alexa$variation[i] == "Black" |  amazon_alexa$variation[i] == "White") {
+     amazon_alexa$product[i] <- "Echo Dot"
+   } else {
+     amazon_alexa$product[i] <- "Echo"
+   } #closing if statement
+ }
>
> alexa_token <- amazon_alexa%>%
+   unnest_tokens(word, text)
>
>
> echo <- alexa_token %>%
+   filter(product == "Echo")
```

```
>
> echo_dot <- alexa_token %>%
+   filter(product == "Echo Dot")
>
> afinn <- alexa_token %>%
+   inner_join(get_sentiments("afinn"))%>%
+   summarise(sentiment=sum(value)) %>%
+   mutate(method="AFINN")
Joining, by = "word"
>
> bing_and_nrc <- bind_rows(
+   alexa_token%>%
+     inner_join(get_sentiments("bing"))%>%
+     mutate(method = "Bing et al."),
+   alexa_token %>%
+     inner_join(get_sentiments("nrc") %>%
+                  filter(sentiment %in% c("positive", "negative"))) %>%
+     mutate(method = "NRC")) %>%
+   count(method,  sentiment) %>%
+   spread(sentiment, n, fill=0) %>%
+   mutate(sentiment = positive-negative)
Joining, by = "word"
Joining, by = "word"
>
> bind_rows(afinn, bing_and_nrc) %>%
+   ggplot(aes(method, sentiment, fill=method))+
+   geom_col(show.legend=FALSE)+
+   facet_wrap(~method, ncol =1, scales= "free_y")
```

```
> ################################################################
> ######## Most common positive and negative words #############
> ################################################################
> library(ggplot2)
>
> echo <- alexa_token %>%
+   filter(product == "Echo")
>
> echo_dot <- alexa_token %>%
+   filter(product == "Echo Dot")
> #Most common positive and negative words for Echo
> bing_counts_echo <- echo %>%
+   inner_join(get_sentiments("bing")) %>%
+   count(word, sentiment, sort=T) %>%
+   ungroup()
Joining, by = "word"
>
> bing_counts_echo %>%
+   group_by(sentiment) %>%
+   top_n(10) %>%
+   ungroup() %>%
+   mutate(word=reorder(word, n)) %>%
+   ggplot(aes(word, n, fill=sentiment)) +
+   geom_col(show.legend = FALSE) +
+   facet_wrap(~sentiment, scales = "free_y")+
+   labs(y="Contribution to sentiment", x=NULL)+
+   coord_flip()
Selecting by n
>
> #Most common positive and negative words for Echo Dot
> bing_counts_echo_dot <- echo_dot %>%
```

```
+   inner_join(get_sentiments("bing")) %>%
+   count(word, sentiment, sort=T) %>%
+   ungroup()
Joining, by = "word"
>
> bing_counts_echo_dot %>%
+   group_by(sentiment) %>%
+   top_n(10) %>%
+   ungroup() %>%
+   mutate(word=reorder(word, n)) %>%
+   ggplot(aes(word, n, fill=sentiment)) +
+   geom_col(show.legend = FALSE) +
+   facet_wrap(~sentiment, scales = "free_y")+
+   labs(y="Contribution to sentiment", x=NULL)+
+   coord_flip()
Selecting by n
> ################################################
> ###### N-grams and tokenizing ###############
> ################################################
>
> library(dplyr)
> library(tidytext)
> library(tidyr)
> library(tidytuesdayR)
>
>
>
> alexa_bigrams <- amazon_alexa %>%
+   unnest_tokens(bigram, text, token = "ngrams", n=2)
>
> library(tidyr)
> bigrams_separated <- alexa_bigrams %>%
+   separate(bigram, c("word1", "word2"), sep = " ")
>
> bigrams_filtered <- bigrams_separated %>%
+   filter(!word1 %in% stop_words$word) %>%
+   filter(!word2 %in% stop_words$word)
>
>
> bigram_counts <- bigrams_filtered %>%
+   count(word1, word2, sort = TRUE)
>
```

```
>
> ########################################################
> ####### VISUALISING THE BIGRAM NETWORK #################
> ########################################################
>
> #install.packages("igraph")
> library(igraph)
> bigram_graph <- bigram_counts %>%
+   filter(n>10) %>%
+   graph_from_data_frame()
Warning message:
In graph_from_data_frame(.) :
  In `d' `NA' elements were replaced with string "NA"
>
> bigram_graph
IGRAPH ee7a2c5 DN-- 60 50 --
+ attr: name (v/c), n (e/n)
+ edges from ee7a2c5 (vertex names):
 [1] echo      ->dot       NA        ->NA        sound     ->quality
 [4] prime     ->day       play      ->music     echo      ->spot
 [7] alarm     ->clock     fire      ->stick     smart     ->home
[10] love      ->alexa     playing   ->music     amazon    ->prime
[13] absolutely->love      amazon    ->music     love      ->love
[16] google    ->home      bluetooth ->speaker   easy      ->set
[19] original  ->echo      echo      ->dots      2nd       ->generation
[22] easy      ->setup     alexa     ->app       34        ->alexa
+ ... omitted several edges
>
> #install.packages("ggraph")
> library(ggraph)
> ggraph(bigram_graph, layout = "fr") +
+   geom_edge_link()+
+   geom_node_point()+
+   geom_node_text(aes(label=name), vjust =1, hjust=1)
>
> ############################################################
> ###### #######
> ################ 4 consecutive words - quadro-gram ########
> ############################################################
> quadrogram <- amazon_alexa %>%
+   unnest_tokens(quadrogram, text, token = "ngrams", n=4) %>%
+   separate(quadrogram, c("word1", "word2", "word3", "word4"), sep=" ") %>%
+   filter(!word1 %in% stop_words$word) %>%
+   filter(!word2 %in% stop_words$word) %>%
+   filter(!word3 %in% stop_words$word) %>%
+   filter(!word4 %in% stop_words$word) %>%
+   na.omit(quadrogram)
>
> ############################################################
> ###### the tf_idf framework  ############
> ######on our quadro-gram #################
> ############################################################
>
>
>
> #####  quadrogram
```

```
>
> quadrogram_united <- quadrogram %>%
+   unite(quadrogram, word1, word2, word3, word4, sep=" ") #we need to unite what we split in the previous section
>
> quadrogram_tf_idf <- quadrogram_united %>%
+   count(variation, quadrogram) %>%
+   bind_tf_idf(quadrogram, variation, n) %>%
+   arrange(desc(tf_idf))
>
```

Citation

- Siddhartha, M. (2018, July 31). Amazon Alexa Reviews. Kaggle. Retrieved December 6, 2021, from https://www.kaggle.com/sid321axn/amazon-alexa-reviews.
- Teague, K. (2020, August 30). 3 Amazon Echo Security features to turn on when you leave the House. CNET. Retrieved December 6, 2021, from https://www.cnet.com/home/smart-home/3-amazon-echo-security-features-to-turn-on-when-you-leave-the-house/.