Characterizing Apple Support tweets using VADER and LDA Topic Modelling.

Sravya Arutla

Spring 2023, 04/26/2023
sarutla@iu.edu

Abstract. Understanding customer feedback is vital for businesses to improve their services. With the rise of social media platforms, companies like Apple Inc. have started using them to provide customer support. Apple offers customer support via Twitter for its range of smart products, including MacBook, iPhone, Apple Watch, Air Pods, iCloud, and Apple Music. In this study, sentiment analysis and LDA topic modelling were conducted using Twitter data with specific hashtags and keywords. The tweets were analyzed to gain insights into the primary technical issues, opinions, preferences, and behaviors of Apple users, as well as people's sentiments towards various Apple products and services over roughly a month's data. This study aims to provide valuable insights into customers' experiences and help improve Apple's overall customer service.

1 PROBLEM STATEMENT

This research study focuses on fetching twitter data using specific hashtags and keywords, pre-processing and analyzing the related tweets to get insights about the major problems, opinions, preferences, and behaviors of Apple device users, as well as people's sentiments on all tweets in during the month of March 2023, regarding various Apple products and services.

Twitter's brevity and ease of use make it an ideal platform for people to share product feedback and user experiences. Companies with a support channel help customers by identifying and resolving issues with their products and services. Sentiment analysis on Twitter tweets can help you understand how Apple Support responds to its customers and how satisfied they are with their products and services. Topic modeling, on the other hand, informs businesses about the most frequently discussed topics regarding their services and products. Both of these methods assist businesses in improving customer support services, identifying product improvement opportunities, and improving overall customer experience.

2 BACKGROUND RESEARCH

There were studies performed on the Sentiment Analysis and Topic modelling on the Twitter data to gain the insights of the overall sentiment present in the data and to know what are the most talked themes from the twitter data. But this study primarily focuses on the tweets extracted from the Apple Support, focusing more on the customer feedback analysis on its products and services. This helps the Apple to understand the patterns in the behaviors if the customers and how could further enhance their customer services and its customer engagement on social media platforms like Twitter, which is mostly used by many people around the world.

3 A PLAN FOR CONDUCTING RESEARCH

3.1 Data Collection

As the Twitter API requires developer access and has rate limits restrictions, the tweets from Twitter were extracted for analysis purposes using the python package "snscrape". This study only looks at tweets that contain the hashtag #AppleSupport and the keyword "Apple Support" in the English language. Approximately 70 tweets per day ranging from March 1st 2023 to March 31st 2023 were collected. In total, 3200 distinct tweets were gathered in order to analyze sentiments and identify the most popular subjects. All these tweets were stored in AppleTweets2.csv file using python. Below are the details of each of the fields of extracted tweets:

Field Name	Description
User	User's twitter handle.
Date Created	Date when the tweet was posted (yyyy-mm-dd hh-mm-ss format)
Number of Likes	Number of likes for the tweet posted
Source of Tweet	Source of the tweet posted from.
Tweet	Comment posted by the users

Table 1: Field descriptions of the extracted data for Appletweets dataset.

```
: ▶| appletweets_df
27]:
```

	User	Date Created	Tweets
0	Stevelee987	2023-02-28 23:58:45+00:00	@AppleSupport any tips to stop my iPhone 13 Pr...
1	AppleSupport	2023-02-28 23:57:11+00:00	@tejbirbedi Thanks for tagging us! We'd be hap...
2	AppleSupport	2023-02-28 23:51:56+00:00	@TheJShield We can help look into this with yo...

Figure 1: Showing the contents of appletweets_df

3.2 Data Analysis

The following steps are included in the Data analysis part.

3.2.1 Data Pre-Processing

In Natural Language Processing, the NLTK python package is used to remove the observed unwanted data. All short words (length = 3) are removed from raw tweets, including twitter handles, URLs, alphanumeric characters, hashtags, contractions, stop words and punctuation. Upper-case letters are all converted to lower case.

Tweets	Cleaned_Tweet
@_benmattison_ Glad we were able to help out. If you need anything in the future, feel free to contact us again.	anything future contact
@applesupport where the ibook g4 released yet	ibook released

Table 2: Raw Tweets vs Cleaned Tweets.

After noise from the raw data has been removed from the cleaned tweet, sentiment analysis of user tweets is then carried out.

3.2.2. Vader Sentiment Analysis

The lexicon- and rule-based sentiment analysis tool VADER (Valence Aware Dictionary and sEntiment Reasoner) is sensitive to both the polarity (positive/negative) and intensity (strength) of emotion in the text. Following pre-processing, the cleaned tweets are fed through the NLTK library into the VADER sentiment analyzer. The polarity scores () of the VADER analyzer divide each tweet's intensity into positive, negative, and neutral scores. Based on these intensity values the compound score is calculated. The compound score is just the normalized total of all the polarity scores, with -1 being the most severe negative and +1 representing the most extreme positive. Following that, the tweet's final polarity is classified as either positive, negative, or neutral, as shown below.

- Positive: (compound score >= 0.05)
- Neutral: (compound score > -0.05) and (compound score < 0.05)
- Negative: (compound score <= -0.05)

In the dataset an additional column is added as Polarity, has these values for the respective tweets.

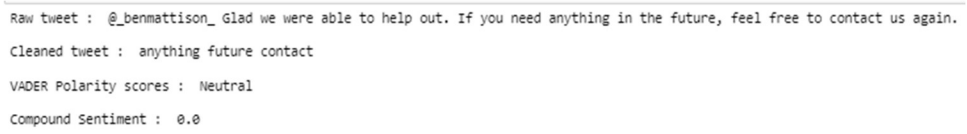

Figure 2: Figure Showing the Sentiment of the tweet using compound score.

By comparing the positive, negative, and neutral polarity scores, rules are generated to classify the sentiment of a tweet as in Figure 2.

3.2.3 LDA Topic Modelling

An unsupervised machine learning algorithm called LDA (Latent Dirichlet Allocation) is used for topic modelling in natural language processing. A corpus of text documents' underlying topics is found using a probabilistic model. The algorithm makes the assumption that each corpus document consists of a fixed number of topics, each of which is represented by a probability distribution over words. The topic-word distributions and the

document-topic distributions are iteratively estimated by the LDA algorithm. The algorithm updates the probabilities that each word belongs to each topic and that each topic is present in each document during each iteration. The most likely topic assignments for each word in each document are created using these probabilities. The 'gensim' library of python is used to implement this, which makes use of a probabilistic generative statistical model to infer particular topics based on terms.

An additional pre-processing of the tweets is done to perform LDA Topic modelling, From the cleaned text obtained from (a)

- The WordNetLemmatizer and Gensim python packages were used, respectively, for tokenization and lemmatization. Lemmatization is the process of reducing words to their root or dictionary form to improve accuracy in text analysis, while tokenization divides paragraphs into meaningful word chunks.
- Bigrams and trigrams creation on the tokienized_tweet from python's gensim is done in order to get more meaning grouping of the topics.
- POS_tagging (Parts of Speech) where Nouns and Adjective are considered is done on bi_tri column of the data frame.
- TF-IDF Vectorization to get the accurate and relevant topics from LDA.

The Document Term Matrix is produced by creating a term dictionary of the bi_tri corpus in which each distinct term is given an index. This document's term matrix is analysed using an LDA model to group related words into a number of groups with high probabilistic value.

4. RESULTS

4.1 VADER Sentiment Analysis

It is seen that, out of 3200 tweets considered, 1729 have positive sentiment, 947 have neutral sentiment, and 524 have negative sentiment. It means that because tweets about Apple Support from users are mostly

positive, we can conclude that Apple Support's customer service is making its users feel good about the services and support it provides. The diagram below depicts a bar chart for negative, neutral, and positive tweets.

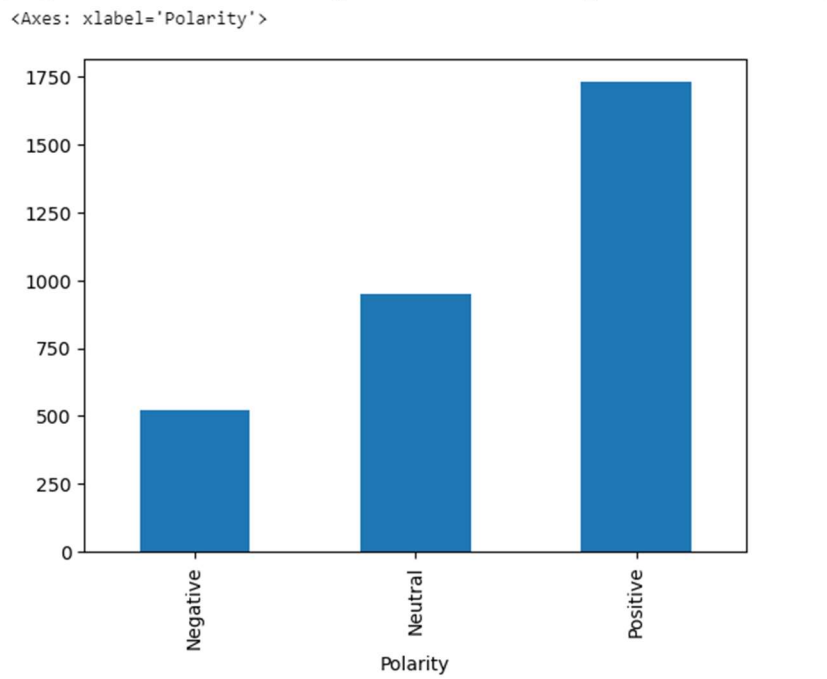

Figure 3: BAR Chart showing sentiments of the appletweets_df.

4.2 LDA Topic Modelling

In total of 5 different topics were chosen for running the model and the tweets are classified into these 5 topics themed as shown below.

	Topic 1	Topic 2	Topic 3	Topic 4	Topic 5
0	apple	solution	thank	thanks	support
1	issue	iphone	iphone	issue	continue
2	step	feedback	apple	happy	offer
3	message	happy	phone	version	hello
4	happy	update	device	hello	iphone
5	detail	device	airpods	information	thing
6	device	option	support	continue	issue
7	thanks	issue	assist	gather	community
8	assist	apple	issue	resolution	twitter_english
9	direct	question	happy	reach	message
10	iphone	right_place	reply	step	apple
11	facetime	thanks	service	device	battery
12	music	suggestion	problem	iphone	version
13	today	great	private	check	device
14	watch	provide	battery	provide	happy
15	start	detail	request	information_determine_proceed	preferred_language
16	continue	version	detail	detail	contact
17	store	service	headed_right_direction	assist	thank
18	question	feature	watch	software	thanks
19	refund	check	macbook	reset	weather

Figure 4: Top 20 words in each topic from the LDA Model.

The analysis and grouping of words in each of the topics can be done as below.

1. ***Topic 1***: The terms "iPhone", "FaceTime", "music", and "watch" imply that the reported problems may be connected to these particular Apple goods and services. The terms "refund" and "store" imply that customers might be having problems with their Apple Store purchases. The word "happy" implies that users might be expressing gratitude for the assistance they received. Users seeking assistance to resolve problems with various Apple products and services appear to be the focus of this LDA topic modelling, which can be referred to as "Apple Product and Service Issues".

2. ***Topic 2***: The keywords "iPhone" and "device" imply that the feedback is about Apple products. The keyword "happy" suggests that users are pleased with the product or service they received. The keywords "right_place" imply that users may be providing feedback in the appropriate platform (for example, Apple's support Twitter account). The keywords "solution" and "improvement" imply that users may be suggesting ways to improve Apple products and services. Overall, the theme of this LDA topic modelling appears to be users providing feedback to Apple about their products and services, including suggestions for improvement and requests for assistance with any related issues, and thus it can be themed as "Apple Product and Service Feedback and Improvement."

3. ***Topic 3***: The keywords "iPhone", "phone", "device", "AirPods", "watch", and "MacBook" imply that the support requests are related to Apple products. The word "battery" suggests that users may be having battery problems with their Apple devices. The keyword "private" implies that some users may be looking for private help with their support requests. Overall, the theme of this LDA topic modelling appears to be related to users seeking Apple customer service and technical support for their Apple products, with a focus on resolving issues related to these products, and thus can be themed as "Apple Customer Service and Technical Support."

4. ***Topic 4***: The keywords "issue", "version", "continue", "step", "device", "iPhone", "information", and "reset" imply that users are having technical problems with Apple devices or software. The keywords "network," "airpods," and "payment" imply that some users may be having problems with connectivity, accessories, or payment. The keyword "information_determine_proceed" implies that users may be providing information to Apple support in order to determine the next steps in resolving technical issues. As a result, this can be categorized as Technical Issues and Troubleshooting.

5. ***Topic 5****:* The keywords "support," "continue," "offer," "hello," "message," "contact," "thank," and "preferred_language" indicate that users are looking for Apple's help and engagement. The keywords "iPhone", "device", "battery", and "version" imply that users may be experiencing technical issues with their Apple devices or software. The term "community" suggests that users may be looking for ways to interact with Apple's customer support community. The keyword "weather" may be an outlier, but it indicates that some users may be looking for help or information about weather-related features on their Apple devices. As a result, this can be classified as a Customer Support and Engagement theme.

The above are the 5 most talked themed for the Apple Support Tweets whose grouping is done manually by analyzing the words and its relation.

Also, the number of topics and contents can be effectively interpreted by using 'pyLDAvis' to analyze the clusters and produce a highly interactive visualization of the data. The size of cluster tells that the topic is most discussed compared to other clusters.

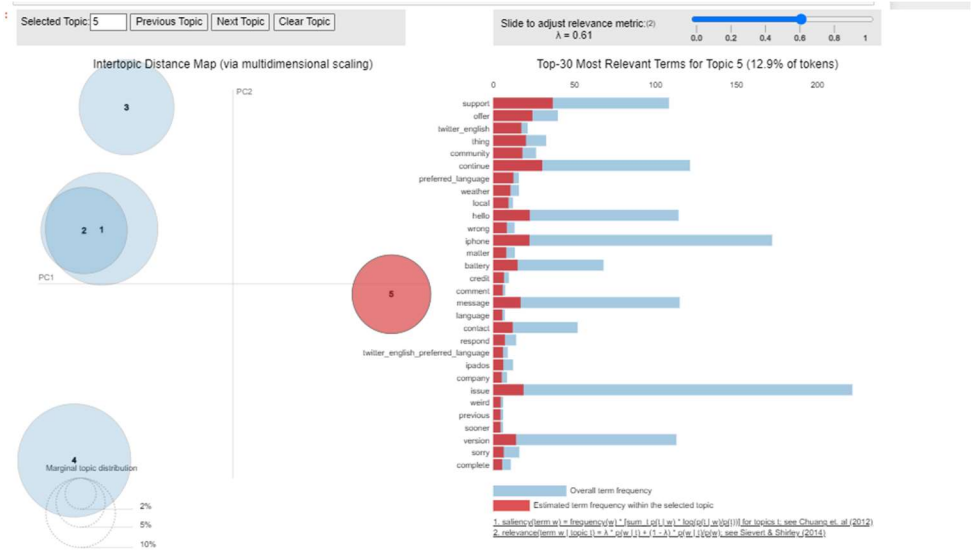

Figure 5: pyLDAvis visualization of topic 5.

5. CONCLUSION

Results obtained from the VADER sentiment analysis and LDA Topic modelling, lets us that during the month of March 2023, most tweets related to the hashtag and Keyword “Apple Support” are positive in sentiment. This tells us that users who are using the platform for the Apple products and services are happy with the quick response and fixes from the Apple Support team.

LDA Topic Modelling results tells us that topic 4 and topic 1 are having approximately the same cluster size as seen from the pyLDAvis visualization of the topics. Hence the most talked themed on Apple Support is Apple Product and Service Issues and Technical Issues and Troubleshooting. The least talked theme could be topic 5 which is Customer Support and Engagement.

Both these results would help the Apple Support to analyze their how the users are responding to the services that are being offered by the

company and what are its most talked themed. This would definitely help their customer feedback analysis.

6. FUTURE WORK

- Taken dataset has only 3200 tweets, if the working dataset is larger, then this would give more accurate results in understanding the sentiments and most talked themes on Apple Support.
- Instead of using the VADER Sentiment analysis, in order to find the exact emotion of the tweets, we could implement ML modelling techniques for getting to know precise emotions from each of the tweets.
- Exploring other topic modelling techniques and comparing its performance with the LDA.

7. LENGTH 2117 Words.

8. REFERENCES

1. Andy Swan August 18. "Apple's Core: Consumer Demand and Sentiment Offer Key Clues Ahead of Keynote - Ticker Tape." *The Ticker Tape*, 18 Aug. 2020, https://tickertape.tdameritrade.com/investing/apple-consumer-demand-social-sentiment-18251.
2. Keita, Zoumana. "Social Media Sentiment Analysis in Python with Vader - No Training Required!" *Medium*, Towards Data Science, 21 Nov. 2022, https://towardsdatascience.com/social-media-sentiment-analysis-in-python-with-vader-no-training-required-4bc6a21e87b8.
3. Li, Susan. "Topic Modeling and Latent Dirichlet Allocation (LDA) in Python." *Medium*, Towards Data Science, 1 June 2018, https://towardsdatascience.com/topic-modeling-and-latent-dirichlet-allocation-in-python-9bf156893c24.
4. "How Twitter Content Powers Your Social Media Marketing Strategy." *Twitter*, Twitter, https://business.twitter.com/en/blog/twitter-content-marketing.html.