

“Key Issue Identification & Emotion Classification of Twitter Reviews for Apple Products and Services (1493 words)”

Pratap Roy Choudhury, MS in Data Science, Spring 2022

Indiana University Bloomington

1. INTRODUCTION

In the present era of technical and digital innovations, Apple Inc. is the giant in the digital and mobile electronics industry. Apple’s various smart devices like MacBook, iPhone, Apple Watch, Air pods and services like iCloud, Apple Music are being used by maximum population and these warrants various technicalities faced by its consumers. Users of these products are using social media networks like Twitter extensively to express their feelings about these products and to report their difficulties for quick resolutions [1]. The most frequently occurring issues are posted by the users that can be identified using topic modeling of the tweets [2]. Through this paper I am going to research about the issues faced by the consumers using the apple products and services, also, what are their reactions during the entire process throughout the year by analyzing the emotions attached to the user tweets [3][4].

2. RESEARCH QUESTION

The study will mainly address the following research question:

- What are the key technical issues and problems faced by the Apple product consumers?
- What are the different emotion trends of the people throughout the year regarding various Apple products and services?

3. METHODOLOGY

3.1. DATA COLLECTION

The data is collected using snsrape between January - December, 2021. As “Tweepy” is unable to fetch historical data, we use the snsrape library of python. We will be focusing on the tweets in English language using hashtags - #AppleSupport, #iPhone, #iWatch, #iCloud, #AppleMusic, #photos, #repair for the data collection. All these hashtags are case insensitive. Majority of the tweets include #applesupport and @applesupport to tag the support team which contain information about the product or services by mentioning the product as keyword. A total of 22048 unique tweets are collected to analyze the emotions and to find the most discussed topics.

Also, a dataset ‘text_emotion.csv’, with already labelled emotions is collected from Kaggle to train our supervised model with the help of which the emotions of apple tweets will be classified.

Below are the details of each of the fields of the emotion labelled dataset:

Field	Description
Author	String that displays the twitter handle or user profile name used in Twitter
Tweet ID	The unique tweet identifier posted by the user
Text	The tweet/comment posted by the user
Sentiment	Emotion labels for each tweet (hate, worry, sadness, hate, love)

Table 1: Data dictionary of the emotion labelled tweets collected from Kaggle

Details of each of the fields of the Apple tweet dataset:

Field	Description
Username	String that displays the twitter handle or user profile name used in Twitter
Datetime	Date when the tweet was posted (yyyy-mm-dd hh-mm-ss format)
Tweet ID	The unique tweet identifier posted by the user
Text	The tweet/comment posted by the user

Table 2: Data dictionary of the tweets collected using snsrape

3.2. DATA ANALYSIS

For topic modeling and emotion detection of the tweets, the texts are needed to be cleaned to remove the unwanted characters and form a meaningful document on which the LDA model and other classification techniques will be applied. The complete analyses steps are explained below.

3.2.1. Text Preprocessing

In Natural Language Processing, the textual data is needed to be cleaned and processed before modeling. With the help of some manual observation, it is found that the tweets contain various characters, symbols, http links, emojis etc. which should be removed. The text pre-processing is done on both the emotion tagged dataset and the apple tweets dataset. The prerequisite steps are as follows -

- I. The texts are transformed to lower-case and then any username (starting with @), # symbols, numbers, URLs, characters that are non-alphanumeric, 'rt' for the re-tweets are removed from the text using regular expression.
- II. Using the stopwords package of NLTK, stopwords like and, or etc., are removed. Words like not, did not, doesn't etc. are not removed as these words can impact the emotions of the text.
- III. For training the classification model, a set of emotions like 'love', 'surprise', 'worry', 'sadness', 'hate' and 'anger' are chosen from the emotion tagged tweets which will be used as training set. The tweets will be classified as one of these emotions as the result.
- IV. The tweets contain many duplicate texts which are dropped. Additionally, emoticons, symbols, pictographs, flags are excluded applying the regex unicode substitution with empty ('') space.

3.2.2. Topic Modeling using LDA method

Bag-of-words information (LDA or TF-IDF) is effective for identifying topics by finding frequent words when texts are coherent within themselves. Here the Latent Dirichlet Allocation (LDA) topic modeling algorithm available in 'gensim' library is used, which utilizes a probabilistic generative statistical model to infer specific topics based on terms. A term dictionary of the cleaned corpus is created where every unique term will be assigned an index, which is then used to create Document Term Matrix. LDA model analyzes this doc-term matrix to cluster into several groups with similar words with high probabilistic value.

The number of topics and contents can be interpreted well by analyzing and creating a highly interactive visualization of the clusters using 'pyLDAvis'. These topics will be the most talked issues or themes by the consumers.

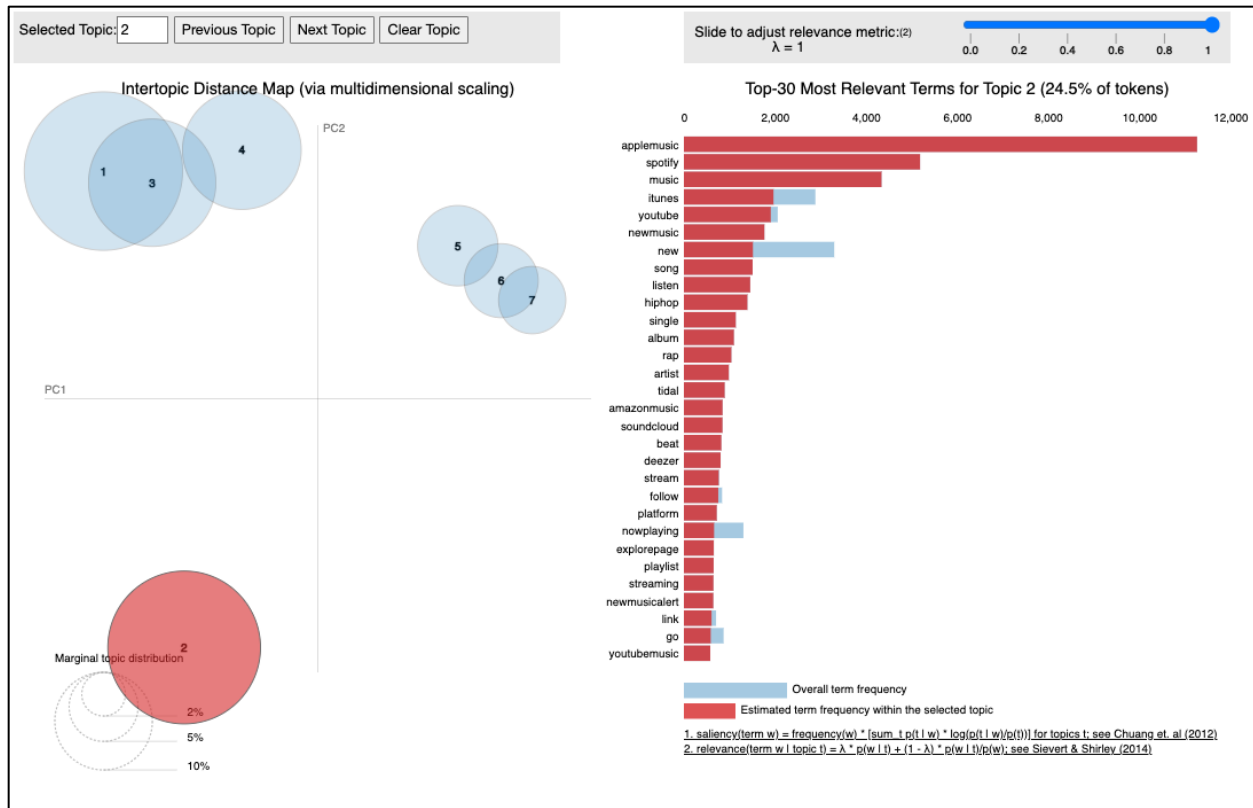


Figure 1: Interactive visualization of topic clusters with frequent tokens

Based on the visualization understanding, model with 5 and 7 topics are created which are manually analyzed to identify the distinguished categories of topics.

3.2.3. Emotion Detection using Machine Learning models

To complete the emotion classification, a supervised modeling technique is followed. The emotion labeled tweeter dataset is for training the classification models. The idea is to perform multiclass classification to classify all the tweet's most popular emotions. The textual data is first converted into vectors using Count Vectorizer and then various predictive modeling techniques viz. Logistic Regression, Decision Tree, SVM, Naïve Bayes and Random Forest are used to train the model. Random Forest is chosen based on its highest accuracy score to classify the customer tweets for products. The cleaned test tweets are then classified as one of the best suited emotions with highest probability score. An emotion trend of the tweets over the year will be generated to understand the user experience and feelings regarding the already launched and upcoming products and services.

Below is the emotion classification of the test set containing apple tweets.

UserName	Tweet Id	Date	Text	cleaned_text	Emotion
iphoneglance	1.356E+18	2021-01-31	What to Choose: iPhone 12 Pro Max or Samsung Galaxy S21 Ultra #samsung #iphone #apple https://t.co/8LKDzBt58s	choose iphone pro max samsung galaxy ultra samsung iphone apple	surprise
ireviews	1.3559E+18	2021-01-31	🔒 LOCK iPhone Apps with PASSCODE or FACE ID on iOS 14 ! #iPhone	lock iphone apps passcode face id ios iphone	worry

			https://t.co/Cc84cGtdVv https://t.co/BcP9m3jJaQ		
Sungyoup_Han	1.3557E+18	2021-01-31	back to #iPhone after using #android for 2 years.. and It feels as if I've never left it at all..	back iphone using android years feels ive never left	love
JMOTA3	1.3657E+18	2021-02-27	@BillGates Says His Preference for @Android Over #iPhone is Due to Pre-Installed Software https://t.co/o0xiXNTwH1	billgates says preference android iphone due preinstalled software	sadness

Table 3: Emotion classified Apple tweets

4. RESULTS

- I. As there are 5 and 7 topics generated with the help of LDA and inspected through pyLDA visualization, a manual observation to the tokens gave a fare idea to construct distinct themes or categories.

Topic 0	Topic 1	Topic 2	Topic 3	Topic 4
['0.037', 'iphone']	['0.013', 'icloud']	['0.085', 'applemusic']	['0.070', 'iphone']	['0.036', 'applemusic']
['0.007', 'photography']	['0.009', 'iphone']	['0.044', 'spotify']	['0.059', 'apple']	['0.018', 'nowplaying']
['0.006', 'bitcoin']	['0.007', 'private']	['0.036', 'music']	['0.025', 'icloud']	['0.017', 'w']
['0.005', 'year']	['0.006', 'launched']	['0.016', 'itunes']	['0.020', 'io']	['0.016', 'brandnewmusic']
['0.005', 'important']	['0.006', 'library']	['0.016', 'youtube']	['0.011', 'ipad']	['0.016', 'standardfm']
['0.005', 'life']	['0.006', 'control']	['0.015', 'newmusic']	['0.009', 'app']	['0.016', 'tuenin']
['0.004', 'screenshot']	['0.006', 'mail']	['0.015', 'new']	['0.008', 'applesupport']	['0.014', 'tuesdayfeeling']
['0.004', 'iphoneplus']	['0.005', 'unlock']	['0.013', 'song']	['0.008', 'new']	['0.014', 'nightmood']
['0.004', 'also']	['0.005', 'redbubble']	['0.012', 'listen']	['0.008', 'phone']	['0.006', 'retweet']
['0.004', 'case']	['0.005', 'wallpaper']	['0.012', 'hiphop']	['0.007', 'android']	['0.006', 'itunes']
['0.004', 'theft']	['0.004', 'east']	['0.010', 'single']	['0.006', 'free']	['0.006', 'make']
['0.004', 'heading']	['0.004', 'overall']	['0.009', 'album']	['0.006', 'user']	['0.005', 'snrtg']
['0.004', 'happiness']	['0.004', 'junk']	['0.009', 'rap']	['0.005', 'get']	['0.005', 'day']
['0.004', 'short']	['0.004', 'beta']	['0.008', 'artist']	['0.005', 'photo']	['0.005', 'global']
['0.004', 'ashesfordreams']	['0.004', 'custom']	['0.008', 'tidal']	['0.004', 'tech']	['0.005', 'addtoplaylist']

Table 4: Top 15 words in 5-topic model

Topic 0	Topic 1	Topic 2	Topic 3	Topic 4	Topic 5	Topic 6
['0.103', 'applemusic']	['0.043', 'iphone']	['0.037', 'applemusic']	['0.016', 'christmas']	['0.096', 'iphone']	['0.059', 'applemusic']	['0.063', 'apple']
['0.039', 'music']	['0.012', 'user']	['0.033', 'nowplaying']	['0.013', 'upcomingartist']	['0.034', 'apple']	['0.044', 'spotify']	['0.057', 'iphone']
['0.037', 'spotify']	['0.012', 'cause']	['0.028', 'w']	['0.013', 'list']	['0.016', 'android']	['0.029', 'youtube']	['0.038', 'icloud']
['0.016', 'new']	['0.010', 'favorite']	['0.027', 'brandnewmusic']	['0.011', 'listening']	['0.014', 'phone']	['0.028', 'music']	['0.026', 'io']
['0.014', 'album']	['0.009', 'photography']	['0.027', 'tuenin']	['0.011', 'window']	['0.012', 'airpods']	['0.023', 'hiphop']	['0.015', 'ipad']
['0.012', 'single']	['0.007', 'art']	['0.027', 'standardfm']	['0.009', 'password']	['0.012', 'case']	['0.022', 'itunes']	['0.013', 'app']
['0.012', 'listen']	['0.007', 'bitcoin']	['0.022', 'tuesdayfeeling']	['0.008', 'private']	['0.011', 'mobile']	['0.021', 'newmusic']	['0.011', 'applesupport']
['0.012', 'itunes']	['0.007', 'global']	['0.022', 'nightmood']	['0.007', 'child']	['0.009', 'tech']	['0.017', 'rap']	['0.008', 'new']
['0.011', 'apple']	['0.007', 'target']	['0.012', 'someone']	['0.007', 'iphone']	['0.008', 'pro']	['0.015', 'artist']	['0.008', 'free']
['0.011', 'stream']	['0.007', 'claim']	['0.008', 'party']	['0.007', 'short']	['0.008', 'technology']	['0.015', 'tidal']	['0.007', 'photo']
['0.010', 'go']	['0.007', 'fake']	['0.007', 'power']	['0.007', 'electronica']	['0.007', 'technews']	['0.014', 'song']	['0.005', 'feature']
['0.010', 'playlist']	['0.007', 'k']	['0.007', 'applemusicplaylist']	['0.007', 'favourite']	['0.007', 'g']	['0.014', 'amazonmusic']	['0.005', 'year']
['0.009', 'song']	['0.006', 'hey']	['0.007', 'fresh']	['0.006', 'refurbished']	['0.007', 'samsung']	['0.014', 'beat']	['0.005', 'get']
['0.008', 'newmusic']	['0.006', 'january']	['0.006', 'friday']	['0.006', 'taylorswift']	['0.007', 'news']	['0.013', 'soundcloud']	['0.005', 'itunes']
['0.007', 'link']	['0.006', 'rule']	['0.006', 'crime']	['0.006', 'add']	['0.006', 'deal']	['0.012', 'deezer']	['0.005', 'user']

Table 5: Top 15 words in 7-topic model

After careful study of the various topics generated and the subjectivity of the tweets, 5 different categories of topics that's speak distinct issues are generated as below.

- a) **Topic 1:** One of the topic is majorly talks about the photography, photos, album, wallpapers etc. in iPhone, iPhone plus devices and there are relevant issues that customers faces during the updates.
- b) **Topic 2:** Another frequent issue that is found relevant to the security, password and passcode, cybersecurity, phone lock and unlock issue, frozen screen, email, repair request to apple-support.

- c) **Topic 3:** People are excited and talked more about the apple music, various music apps like amazon music, spotify, iTunes, songs and artists, music albums and playlist, music genre, streaming and various other services related to music store.
 - d) **Topic 4:** Another topic that is categorised related to the iPhone and iPad, Appstore, games and apps and their subscription related queries.
 - e) **Topic 5:** An interesting topic is found that talks more about technologies and different OS - iOS and Android, phone models, comparison with Samsung devices, accessories, charger issue, refurbished model and sale, tech news etc.
- II. Among all the classified emotions, people are more ‘worried’ because of the various updates, subscriptions, device break-down and technical issues that frequently occur in Apple devices. 62% of the users are worried, whereas 16% are ‘sad’ and 14% customers or potential buyers express ‘love’ or excitement for the products and services. Very less population are really ‘hate’ or express extreme disgust as we can see only a 2% of them under this emotion.

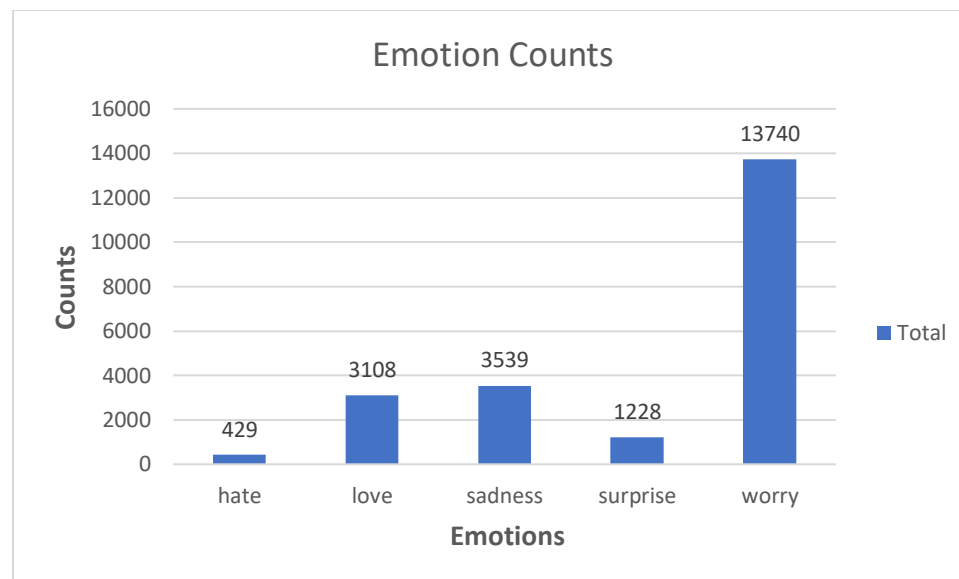


Figure 2: Emotion counts of the collected between 2021-01-01 to 2021-12-31

- III. As the emotion trend shows, there is a spike in ‘worry’ during the months of June and August. People are also ‘surprised’ in September, they express ‘love’ from June to September. These all are the indication of the new product launch which usually occur during September every year, when Apple launches new iPhone and iPad models and people are always excited about it. It can also be seen that the ‘sadness’ has a spike during May and ‘worry’ is typically high during January till May, as Apple pushes new updates in all devices and software that cause several issues and problems to the users. A drop in ‘love’ emotion on October indicates that certain section of the users are not liking the new product.

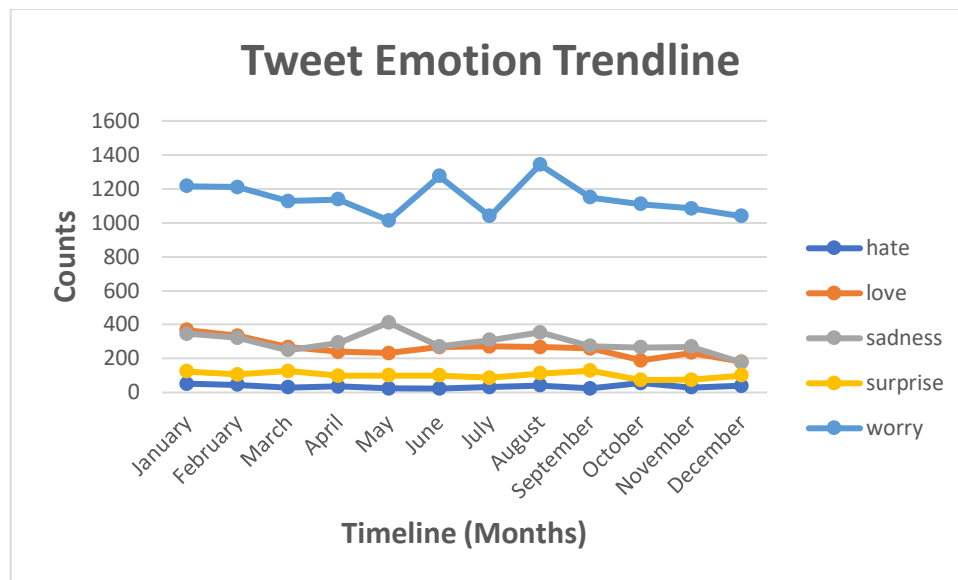


Figure 3: Emotion trend of users over a period of one year - 2021

- IV. The tweets are also categorised under different product and services to understand which products or services have the big buzz. Majority of the users are concerned for iPhone related issues, whereas significant amount of the population commented about iCloud and storage, Apple Music, iTunes and MacBook. These are the sections which should be taken care of by Apple to resolve customer issues and queries.

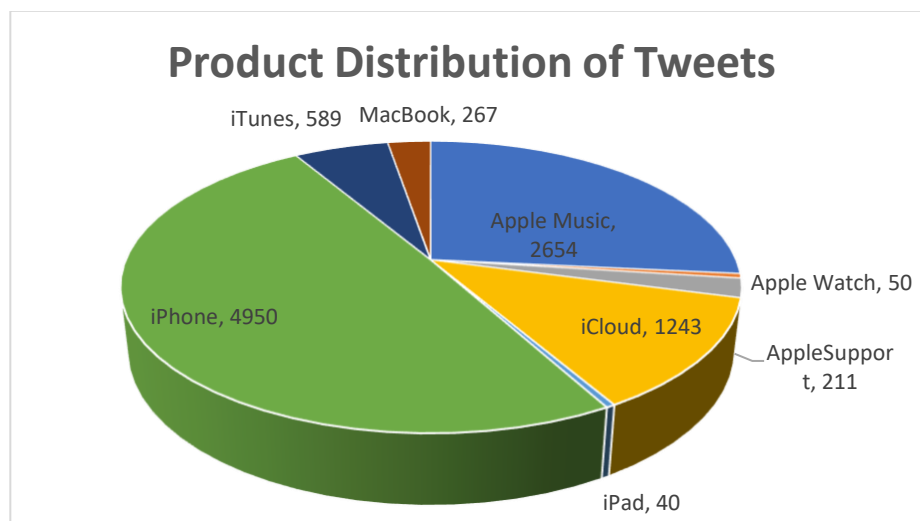


Figure 4: Tweet distribution in major product categories

5. CONCLUSION AND LIMITATIONS

The study intends to find out the major topics and categories regarding which the Apple product users shared their concerns and to find out the emotion of the users. The topic modeling technique can identify five different categories about which people are talking the most that includes iPhone and iPad and

their applications, music, password and security concerns, device, and accessories etc. Apart from that, the emotion detection technique identifies various emotions attached to the tweets.

There are obvious limitations to the research, as the data is collected for 2021 when not much of the products were launched. The data are collected at an overall product level and that may be biased towards a certain product type or issues. As the data is based on 2021 which is post covid period, the emotion of the population can also be biased as there are certain limitations to the amount of data collected. There is still a scope to analyze the tweets if collected geo-location wise.

REFERENCES

- [1] Tilak Naik, G. & Vaikunth Pai T. (2018). A Study on Products and Services of Apple Inc. *International Journal of Case Studies in Business, IT and Education (IJCSBE)*, 2(2), 67-76. DOI: <https://doi.org/10.5281/zenodo.1491388>
- [2] Peter J. Franz, AM, Erik C. Nook, AM, Patrick Mair, PHD and Matthew K. Nock, PHD. "Using Topic Modeling to Detect and Describe Self-Injurious and Related Content on a Large-Scale Digital Platform". *Suicide and Life-Threatening Behavior* 50 (1) February 2020 5 © 2019 The American Association of Suicidology DOI: 10.1111/sltb.12569
- [3] Brynielsson, J., Johansson, F., Jonsson, C. et al. Emotion classification of social media posts for estimating people's reactions to communicated alert messages during crises. *Secur Inform* 3, 7 (2014). <https://doi.org/10.1186/s13388-014-0007-3>
- [4] B. V. Namrutha Sridhar, K. Mrinalini and P. Vijayalakshmi, "Data Annotation and Multi-Emotion Classification for Social Media Text," *2020 International Conference on Communication and Signal Processing (ICCSP)*, 2020, pp. 1011-1015, <https://ieeexplore.ieee.org/document/9182362>