# Group 11 Research Report:
# Roku User Comments Analysis

Shengkang Zhou
Heran Bai
Shucen Yu
Shauna Han
Yachen Xiao
Megan Dandolu Reddy

## Abstract

Roku is a smart TV that uses technology from a streaming media player. The first Roku model was developed in collaboration with Netflix, and since then Roku has gained enormous popularity, having more than 55 million active subscribers in 2021. In this group project, we investigate the popular topics and sentiment change of Roku users on Amazon and Twitter for versions OS 9.4, OS 10, and OS 10.5. We aim to generate findings using LDA topic modeling to extract popular topics from amazon reviews, and sentiment analysis to measure users' opinion on Roku. Through our analysis of product reviews, we hope to identify what users like and dislike about their Roku products, and offer some suggestions for the next product development.

We hypothesized that as the model version updates over time, there will be a positive correlation between the upgraded OS version and having more positive reviews. Our findings from LDA Topic Modeling and Sentiment Analysis aligned with our hypothesis that Roku's frequent software updates contributed to an enhanced user experience. We realize that our results were affected by multiple factors, such as using bounded time intervals to gather data, and we acknowledge that limitations in this study need to be addressed further.

## Introduction

Online reviews are indispensable to digital marketing in the era of social media, as they are great avenues for grabbing consumer's attention and for publicizing businesses.

Consumers do not have the opportunity to engage directly with brand products, as many of them prefer online shopping over in-person shopping. It has become a norm to search for products online before making any purchases. As such, online reputation is very crucial with its power to enhance or devalue brand image. Knowing what consumers like or dislike is the key to successful marketing campaigns. In this project, we aim to examine Roku products based on online reviews from Amazon as well as tweets from Twitter. Using NLP techniques, we hope to better understand how Roku products are perceived by consumers and to make business recommendations which yield tangible outcomes.

Roku is a device that streams media (shows, movies, and even music) from the internet to TV. It offers a practical and affordable way to add internet streaming or expand internet streaming options to TV and home theater viewing experiences.

In this group project, we analyze the sentiments of Roku users for each upgraded version: OS 9.4, OS 10, and OS 10.5, with  three different analyses: LDA topic modeling, sentiment analysis, and text analysis. Firstly, we use the LDA topic model to identify popular topics from Amazon reviews and Tweets data and to capture key words in main topics. To better understand users' reviews for each version of Roku products we conduct sentiment analysis by calculating the sentiment score for each comment. Then, we generate the word cloud and sentiment visualizations to determine the logic and relationships of the words.

Specifically, we analyze customers’ sentiments and emotions using tweets and amazon customer reviews. This is because we believe that these are two social platforms that people will refer to before purchasing an item. Through our analysis of product reviews, we hope to identify what users like and dislike about their Roku products, such as whether the version upgrade has brought a better experience, what other problems persist with the products, what aspects of the products still need to be improved, and how people perceive the products in comparison with those of competitors. We also look at the official updated documents from Roku’s website to compare with our findings. We want to discover whether the changes noted in these documents are also being echoed by the users.

**Data Preparation**

To answer whether the operation system updates of Roku Ultra from 9.4 to 10.5 improve user experience, we decided to use the data from two sources: Amazon and Twitter. We assumed that the comments correspond to the most recent OS version when they were written.

For Amazon, we extracted the user reviews of Roku Ultra from September 2019, when this product was first introduced, to February of 2022, before the release of OS 11.0. To extract the data, we generated API token and keys from the Rainforest API, and set no other filters except for time. In total, we extracted 1,870 reviews. However, for the three OS versions, the sample number was unbalanced: OS 9.4 had 865, OS 10 had 543, and OS 10.5 had 462 reviews. This unbalanced sample size might affect further analysis. The final dataset contains the following information: 1) the title of the review, 2) the body of the review, 3) the rating on a scale of 5, 4) the time the review was written and 5) the OS version.

To get the data from the Twitter API, we made a developer account for an academic purpose on developer.twitter.com website. From the developer platform, we obtained API keys and tokens for the use of Tweepy, an open-source Python package that allows us to collect a set of metadata (tweets). Utilizing Tweepy in our python code, we could scrape 500 tweets each time. 500 tweets constituted about three months' duration of tweets, and we downloaded tweets multiple times to obtain a six months data for each version and merged them into one csv file. We filtered tweets that were from the US. In the data we selected only ‘twitter Id’ and ‘text’ columns for our analysis.

After gathering raw data from the two platforms, we performed the preprocessing techniques to clean the body of each document. First, we cleaned the documents by applying regex grammar to remove numbers, punctuations and special characters. Then we removed all the comments in a non-English language by comparing each word in the comments with NLTK library’s English dictionary. Every word that was not matched was dropped. Since the first two steps truncate the raw documents, we need to make sure the surviving ones are long enough to be informative. Therefore, we dropped every comment that had less than 15 words. After that, we adjusted the stop word dictionary to include words that we want to treat as stop words, like “ultra”, “amazon” and “roku”. Applying a new stopword dictionary helped us focus more on the remaining texts that appeared with similar term frequencies and ensured a better topic modeling. Lastly, we stemmed the documents using the PorterStemmer API.

The last step in data preparation was to separate the documents into respective platforms and OS versions, where we generated 6 tables (3 OS versions for 2 platforms). We further combined the Amazon and Twitter data to produce 3 new combined dataframes for the corresponding 3 versions.

## Methodology

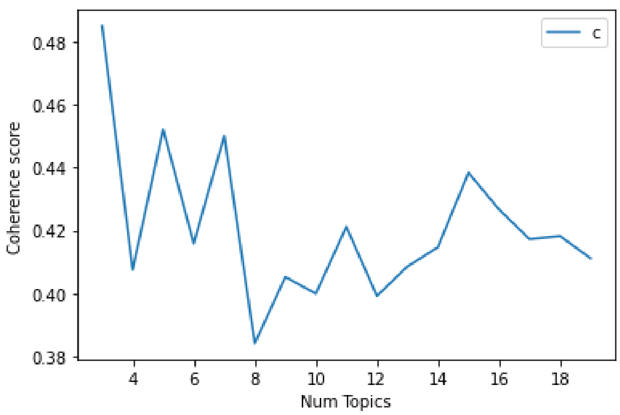
### 1. LDA Modeling

LDA is a topic modeling technique used to identify hidden relationships in text data by classfying texts in a document to particular topics. We performed LDA modeling on our pre-processed datasets of Amazon reviews and tweets to organize our unstructured texts and to identify topics (prominent key words).

In our first step, we performed additional text cleaning to our preprocessed dataset by removing emails, newline characters, and quotations. Then we created bigram and trigram models to consider bigrams and trigrams that occur together as one word, and defined a function to lemmatize the texts in our dataset.

After we further pre-processed our datasets, we created the dictionary and corpus needed for topic modeling. Corpus is a document-term matrix, and we built it with the Gensim's Dictionary function (doc2bow).

To decide the optimal number of topics for our LDA model, we computed coherence scores for the number of topics in the range from 3 to 20. For instance, when we performed the LDA modeling on Amazon reviews for Roku version 10.5, we obtained different coherence scores for each number of topics in our predefined range [fig 1]; we achieved the highest coherence score of 0.4053 when the number of topics was 4.

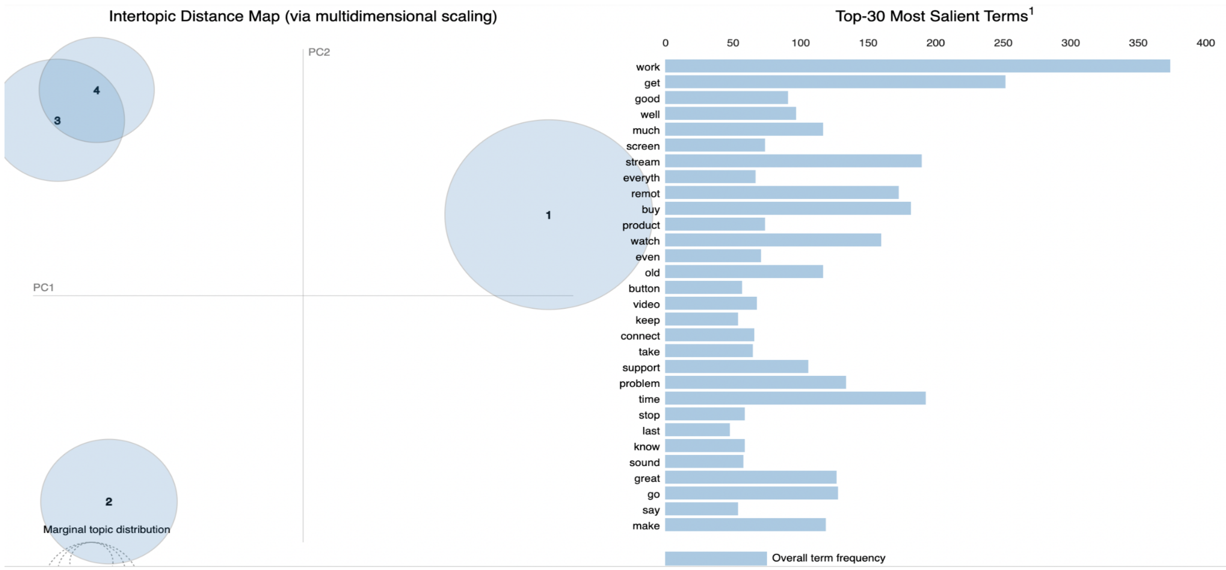



[fig1]

```
(0, '0.027*"remot" + 0.011*"work" + 0.008*"well" + 0.008*"wifi"')
(1, '0.018*"stream" + 0.014*"work" + 0.012*"great" + 0.010*"cabl"')
(2, '0.013*"connect" + 0.010*"support" + 0.009*"got" + 0.009*"work"')
(3, '0.013*"remot" + 0.011*"app" + 0.009*"work" + 0.009*"new"')
```

[fig 2]

We used the LDA to observe key words in dominant topics. In the output above [Fig 2], each topic showed the combination of key terms appearing within the topic; these terms are the words that appeared with highest probability. For example, the words “remot(e)” , “work”, “well”, and “wifi” appeared in Topic 1, 0.027 (2.7 %), 0.011 (1.1%), 0.008 (0.8%), and 0.008 (0.8%) respectively.




[fig 3]

As the last step of the LDA modeling, we visualized the topics using a web-based interactive map, pyLDAvis. Each bubble on the left side of [fig 3] represents topics. The larger the bubble, the more dominant the topic is in the dataset; each text in the dataset is composed of one or more topics. The right side of the interactive map is the top 30 words of the dataset; estimated term frequencies for each word are indicated in the map.

| | **OS 9.4** | **OS 10.0** | **OS 10.5** |
|---|---|---|---|
| **T1** | stream/pictur/support/atmo | watch/stream/channel/live | watch/steam/channel/live |
| **T2** | remot/stream/replac/unit | get/remot/love/stream | remot/great/day/good |
| **T3** | app/remot/watch/fire | work/like/remot/app | stream/devic/connect/issu |
| **T4** | | work/stream/get/remot | |

| | | | |
|---|---|---|---|
| **T5** | | remot/stream/watch/apl | |

[Table 1]

We observed an interesting correlation in the data: the LDA results in [Table 1] showed that dominant terms in topics had a positive connotation such as “remot(e)”, “like”, “love”, “good”, and “stream”. Additionally, many of the top 30 salient keywords in [fig 3] were also positive, such as “work”, “get”, “buy”, “support”, “great”. The LDA result coincided with our hypothesis that users had a positive experience with the Roku Product, and that with each software upgrade, the product received better reviews.

2. **Sentiment Analysis**

To better understand the topics generated by the LDA model, and to capture the overall trend of the reviews, we decided to use sentiment analysis by calculating the sentiment score for the comments.

The positive and negative word list we chose is from the *Hu and Bing 2004* word list. To match the stemmed version of words in the comments, we also stemmed the words in the *Hu and Bing 2004* word list. The algorithm uses value p to keep track of how many words in a comment match a word in the positive word list, and n for the negative. The overall sentiment score is calculated by the following formula:

***sentiment=(p-n)/(p+n)***

By dividing the total occurrence of the sentiment words, the sentiment score is normalized between -1 and 1, with -1 meaning totally negative, 0 meaning neutral and 1 meaning totally positive. For comments without any match, they receive a score of 0 and are treated as neutral.

After we generated sentiment scores for each comment, we first investigated the overall trend of sentiment by comparing the mean score over the updates for Amazon, Twitter and the combined dataset. As [Table 2] suggests, all sentiment scores were greater than 0, but as each new update takes place, the sentiment score dropped. To examine whether the drop was statistically significant, we performed the one-way ANOVA (alpha=0.05, two-tailed) and received significant results.

| | Amazon | Twitter | Combined |
|---|---|---|---|
| OS 9.4 | 0.5532 | 0.4282 | 0.4705 |
| OS 10 | 0.5504 | 0.3934 | 0.4542 |
| OS 10.5 | 0.4742 | 0.3829 | 0.4103 |
| p-value | 0.01 | 0.02 | 0.01 |

[Table 2]

The second test was to investigate whether certain keywords that popped up in the LDA result related to more negative sentiments. We selected “connect” and “battery” as keywords from the LDA topics, and postulated that when a comment contained these words, it was more likely to talk about problems and issues, rather than praising the product. Therefore, we separated documents into two sets by whether they contained the keywords, and compared their mean sentiment score.

| review with *“connect”* | review without *“connect”* |
|---|---|
| 0.4210 | 0.4890 |

[Table 3]

| review with *“battery”* | review without *“battery”* |
|---|---|
| 0.3581 | 0.4857 |

[Table 4]

The results shown by [Table 3] and [Table 4] confirmed our hypothesis.

### 3. Word Cloud Visualization and Sentiment Visualization

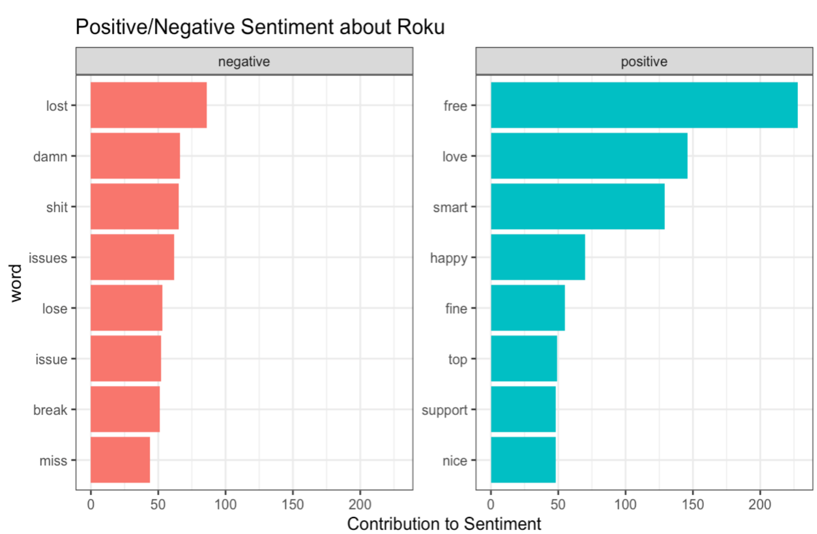
We generated a total of six word clouds [fig 4 and 5] using Python, and this was to compare keywords from two data sources, Twitter and Amazon reviews for each version of Roku product. In Python, we used matplotlib, pandas, and word cloud libraries, and each word in the word cloud indicated its frequency in the dataset.


[fig 4; word clouds using tweets for version OS 9.4, 10.0, 10.5 respectively]


[fig 5; word clouds using Amazon reviews for version OS 9.4, 10.0, 10.5 respectively]

We found that similar keywords related to product functionality such as “remote”, “channel”, and “device”, which appeared in all three versions of Roku. A key difference between word clouds of the tweets and those of the Amazon reviews was that company or brand names (ie. Apple, Youtube, Amazon, Google, etc) appeared more noticeably in the word clouds with tweets. Possible reasons can be postulated to: 1) Twitter being a public and open platform to share comments by any user, 2) Twitter’s easier user interface design to write a comment compared to Amazon reviews, and 3) the possible regulation of online reviews. While the word clouds offered a comprehensive graphic and enabled us to better understand the data, it may not have been the best tool when it came to the accuracy of comparison. We pre-processed the datasets of tweets and amazon reviews separately, and there were slight differences in the text cleaning and stemming steps which may have caused some inconsistency in the analysis.

Next, we performed a sentiment analysis using R and the Bing lexicon. We counted the number of unique words found in tweets, then joined the words extracted from tweets with the sentiment data from Bing. We chose to use the Bing lexicon, as it categorizes words in a binary method into positive and negative categories. We combined tweets for all three Roku versions, and obtained sentiment analysis as in the illustration in [fig 6]. We found that there were more positive tweets than negative tweets in our dataset. For instance, there were no words in the negative sentiment category that had more than 100 times of appearance in the dataset. In contrast, the positively categorized words such as “free”, “love”, “smart” appeared more than 100 times in our dataset, substantiating our hypothesis that Roku products are perceived positively by users.




[fig 6]

**Results**

We used the key words generated by LDA to compare with the update summary posted by the Roku official website. We investigated whether the software update contributed to enhanced user experience, or there was still room for improvement. The results of comments have been divided into three categories: ‘Common features with frequent updates’, ‘System and Hardware Connection Upgrades’ and ’New features and Contents’. The table below concludes what we have found for the three categories for all data sources ( Amazon, Twitter and Combined).

- Common features with frequent updates

| Version | Amazon | Twitter | Combined |
|---|---|---|---|
| **OS 9.4** | • Roku device with their voice remotely using the Home app and Siri on iPhone, iPad, Mac, Apple Watch, or HomePod | • All Roku devices provide easy access to watch free TV, live news, sports, movies, and more. | • Not Founded |
| **OS 10.0** | • Live TV offers easy access to cable alternatives, including Hulu + Live TV, fuboTV, Philo, Sling, and YouTube TV | • All Roku devices provide easy access to watch free TV, live news, sports, movies, and more | • Not Founded |

| | | | |
|---|---|---|---|
| **OS 10.5** | • The Home tab provides easy access to the latest entertainment and channels, and offers a new way to explore Zones | • More access to streaming live TV (U.S.) | • Roku TV users have quick and easy access to live TV<br>• Roku Voice Remote Pro owners will also get tips on using hands-free voice |

- System and Hardware Connection Upgrade

| **Version** | **Amazon** | **Twitter** | **Combined** |
|---|---|---|---|
| **OS 9.4** | • Many people have the issue of the remote doesn't work even if they reset the device, so they have to open the app to control ROKU, which is annoying | • Roku TVs let you access free live broadcast TV using your home's antenna | • Often compared with apple TV |
| **OS 10.0** | • The earbud option of the remote uses up battery really quick | • Not Founded | • Not Founded |
| **OS 10.5** | • Ruku able to combined in many ways to best suit your home setup speaker | • Improvements to the Roku mobile app<br>• Users will experience faster channel launch. | • Not Founded |

- New features and contents

| Version | Amazon | Twitter | Combined |
|---|---|---|---|
| **OS 9.4** | Often compared with fire TV stick | Often compared with fire TV stick | Often compared with fire TV stick |
| **OS 10.0** | Support better Wifi connection<br>Rich and immersive sound experience with the new Virtual Surround setting | Consumer Bought it to watch the nanny<br>Join the club to watch a show | Not Founded |
| **OS 10.5** | Automatic Wi-Fi® network detection technology ensure smooth streaming experiences for Roku users<br>Brings exciting new interactive features, full surround sound capability | Live TV also offers easy access to cable alternatives, including Hulu + Live TV, fuboTV, Philo, Sling, and YouTube TV | Not Founded |

According to the Amazon review analysis, the keys for Roku's success can be attributed to: easy access to different cable alternatives and shows, automatic switch WIFI to provide better streaming experiences, and voice control functions and good sound system experience. Beside these, we also found some unsolved issues, including complaints about the remote control not working even when the user reset the device, and the short-lived battery for the remote earbuds option. From the Twitter comments, we found out that Roku provides easy access to different TV shows, live shows and movies;people purchase Roku because they want to enjoy its exclusive contents. Lastly, the comments from the combined data show that people often compared other brand products with Roku, and they were satisfied with the Roku’s quick access and voice control functions.

Our findings are parallel with our assumption that Roku’s frequent software updates contributed to an enhanced user experience. The overall comments were mostly positive although there were some of the key words including “issues”, “break”, and “miss”. According to our sentiment analysis, while each OS update resulted in a decreased in the sentiment score, the overall score was positive, which again confirms our assumption that software updates contributed to an enhanced user experience.

**Business Applications**

The approaches we used for this project gives us a basic understanding of the social feedback from the two major social platforms. In total, three applications were generated to provide meaningful thoughts for the company.

The first application is sentiment monitoring. Our project provides a general idea on how the customers or the netizens feel about the product or the company. For example, if the sentiment score fluctuates at a certain range for a long period of time, and the score suddenly jumps out from that range, the company should be alarmed. The company needs to find out whether something has gone wrong with the product, or some market opportunity has popped up awaiting to be seized.

The second implication is the pre-stage social feedback. Examining every comment or every tweet can be time-consuming, but with models such as LDA and sentiment score analysis, the process would be much easier. The product manager would immediately grasp what topics are frequently mentioned, and then check comments with such topics manually to know what exactly the problems are. For our project, we found out that the connect issue and the battery issue are the topics that were frequently mentioned, and perhaps the managers in Roku should take a look.

The third implication is precise advertisement. There are certain patterns in different platforms. As we observed, people who buy Roku on Amazon are more sensitive to the quality of the product, while twitter users love to compare the product with other major competitors such as Fire TV and Apple TV. Companies can take advantage of this finding and create tailored contents for users on different platforms. More research can be done here to find out what type of content serves best for each platform.

**Limitation**

One limitation of our research was the bounded timeframe used in our datasets. As we were obtaining real time data, we had to set time bounds to download the data from Amazon's Rainforest API and Twitter API. The specific dates we chose for downloading data were purely arbitrary. For future projects that use real time data, we could predefine time bounds more consistently across each download.

Another possible limitation of the topic modeling using LDA is that it is difficult to tell if it is working correctly as the number of topics are set by the user, making the resulting distribution of topics surfaced from LDA less accurate due to subjectivity. Certain keywords surfaced in our LDA and although we hypothesized that these words were associated with negative reviews based on the sentiment analysis, we cannot be sure that this was the case.

A third possible limitation of sentiment analysis is its inability to detect sarcasm or slang. If a customer uses sarcasm with negative words in a positive review, for example, the sentiment of this review may still be low even though the review is actually positive.

**Conclusion**

In this project, we applied the LDA model, sentiment analysis and word cloud visualization to the Amazon and Tweets data from September of 2020 to February of 2022. Our findings meet our expectation that some words such as " streaming", "free", "TV", and "remote" will surface as the main topic, but we also noticed that topics regarding the system and hardware update and new features would appear in the main topics, and this can provide useful information for the company to know about their problems and optimize their business strategies accordingly. The result is rich in information. With more time and more data on this NLP work flow, there is definitely room for improvement of this project.