Research Report:
Roku User Comments Analysis

Abstract

Roku, a streaming media player-based smart TV, has gained widespread popularity with over 55 million active subscribers as of 2021. This project aims to analyze the topics and sentiment of Roku users on Amazon and Twitter for OS versions 9.4, 10, and 10.5. We will use LDA topic modeling to identify popular topics from Amazon reviews and sentiment analysis to assess users' opinions on Roku. The objective is to understand what features users appreciate and criticize about their Roku products and provide recommendations for future product development.

Our hypothesis stated that the upgrade of OS versions would lead to a higher proportion of positive reviews from users. Our findings, obtained from LDA topic modeling and sentiment analysis, affirm this hypothesis and indicate that frequent software updates improve the user experience. Nevertheless, it is crucial to consider potential limitations such as the restricted data collection period and the subjective nature of setting the number of LDA topics, and further research is required to address these limitations.

Introduction

Online reviews play a crucial role in digital marketing and shaping a brand's image in the era of social media. With the rise of online shopping, consumers often rely on these reviews before making a purchase, making a brand's online reputation critical to its image. To effectively market its products, it's essential to understand consumer preferences and opinions. This project seeks to understand consumer perceptions of Roku products through analysis of Amazon reviews and Twitter data using NLP techniques. Our objective is to gain insights into customer sentiment towards the upgraded versions of Roku's operating system (OS 9.4, OS 10, and OS 10.5).

We will use LDA topic modeling, sentiment analysis, and text analysis to uncover popular topics, sentiment scores, and relationships between words and sentiments. Our focus on Amazon reviews and Twitter is driven by the belief that these are two platforms that consumers frequently consult before making a purchase. Our analysis will identify users' likes, dislikes, experiences, and perceptions of Roku products compared to competitors. We will also compare our findings with official updates from Roku's website to determine if there is alignment.

Data Preparation

To assess the impact of the upgrade from OS 9.4 to 10.5 on the user experience of Roku Ultra, we analyzed data collected from two sources: Amazon and Twitter. To obtain the Amazon data, we used the Rainforest API to gather 1,870 reviews of the Roku Ultra from September 2019 to February 2022, prior to the release of OS 11.0. The final dataset included review title, review text, rating, date, and OS version. The sample size was unbalanced, with 865 reviews for OS 9.4, 543 for OS 10, and 462 for OS 10.5, which could affect the analysis results.

For Twitter data, we created a developer account on developer.twitter.com and used API keys and tokens to collect tweets with the Tweepy library in Python. The tweets were filtered to only include those from the US and were collected over a six-month period for each OS version. The final Twitter dataset included the Twitter ID and text of the tweets.

We applied preprocessing techniques to clean and prepare the data for analysis, including removing numbers, punctuation, and special characters using regex, filtering out non-English comments by checking each word against an English dictionary, and removing short comments (less than 15 words). We also updated the stopword dictionary to include desired stopwords and stemmed the remaining documents using the PorterStemmer API. Finally, we divided the cleaned data into separate tables for each platform and OS version, resulting in 6 tables (3 OS versions for 2 platforms), and combined the Amazon and Twitter data for each of the 3 OS versions to create 3 new combined datasets.

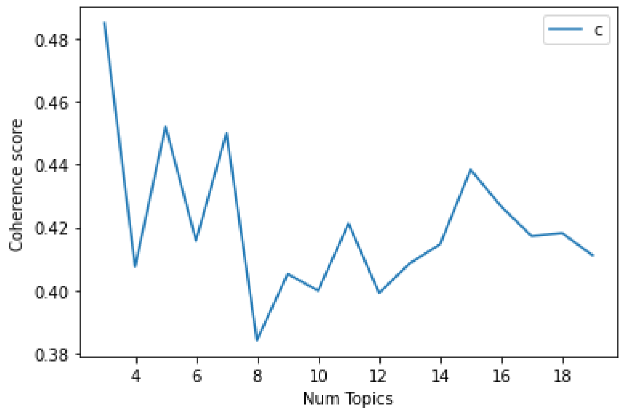
Methodology

1. *LDA Modeling*

LDA (Latent Dirichlet Allocation) is a text mining technique that groups similar texts into topics. We applied LDA to Amazon reviews and tweets to uncover the main topics and keywords in the data.

First, we cleaned the data by removing emails, newline characters, and quotations. We then created bigram and trigram models to consider word combinations as one word and lemmatized the texts. We used the Gensim's Dictionary function to build a document-term matrix and computed coherence scores to determine the optimal number of topics for our LDA model. We obtained different coherence scores for each number of topics in our predefined range [Fig 1]; the highest coherence score of 0.4053 was achieved when the number of topics was 4.

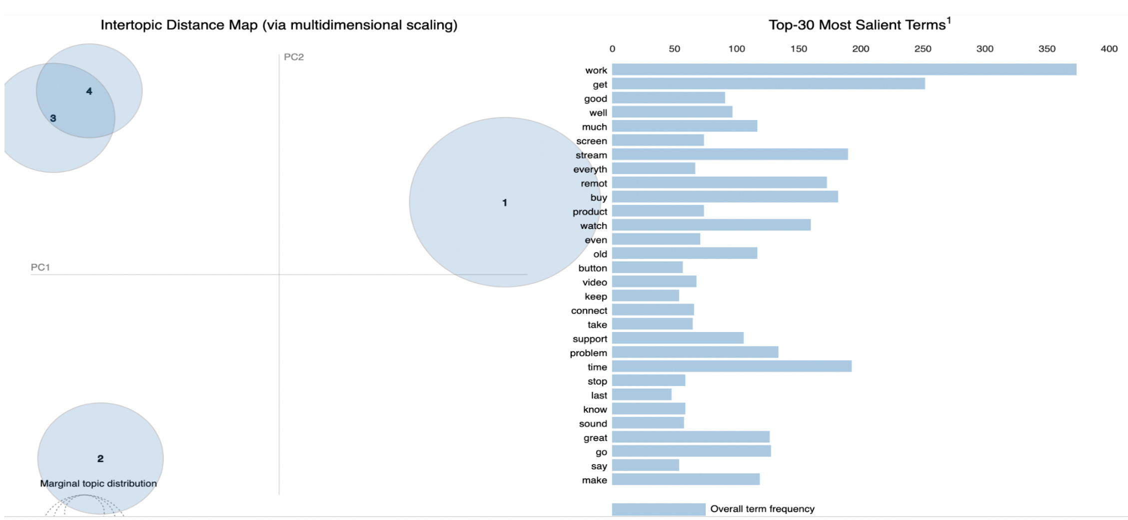
The results of LDA showed the dominant terms and key words in each topic [Fig 2]. The analysis of the interactive map in pyLDAvis revealed the main topics and sentiments expressed in the reviews and tweets. The correlation between positive keywords and dominant terms suggests that users had a positive experience with Roku and that the product received better reviews with each software upgrade [Table 1]. [Fig 3] displayed the result of LDA analysis with a list of the top 30 prominent keywords, many of which had a positive connotation.



[Fig1: Coherence Score by Number of Topics]

```
(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: Probability of key terms appearing in each topic]



[Fig 3: The result of LDA analysis; top terms for each topic in a visually appealing manner]

This visualization enables us to effortlessly recognize the key themes that dominate in the data and the terms that define each theme. By examining the pyLDAvis interactive map, we were able to uncover the major topics and keywords in the Amazon reviews and tweets for each version of the platforms. This gave us a deeper understanding of the sentiments conveyed in the comments, which can be utilized to enhance the performance of the operating system.

	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: the result of LDA for each OS version]

2. *Sentiment Analysis*

To gauge the sentiment of the reviews, we conducted sentiment analysis to calculate the sentiment scores for each comment. We used the Hu and Bing 2004 positive and negative word list, stemming the words to align with the stemmed words in the comments. The algorithm tracked the number of words in a comment that matched a word in the positive list (p) and the number of words that matched a word in the negative list (n), and calculated the sentiment score using the formula:

$$\textbf{sentiment}=(p-n)/(p+n)$$

The score ranges from -1 to 1, where -1 represents a negative sentiment, 0 represents a neutral sentiment, and 1 represents a positive sentiment. If a comment did not contain any matching words, it received a score of 0 and was considered neutral.

We compared the mean sentiment scores for Amazon, Twitter, and the combined data over each update. As shown in Table 2, all sentiment scores were greater than 0, but decreased with each update. To determine if this decrease was statistically significant, we performed a one-way ANOVA (alpha=0.05, two-tailed) and obtained 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: Sentiment score for different data sources]

We found that the sentiment scores for comments containing the keywords "connect" and "battery" were lower compared to those without these keywords. We validated this result with a t-test ($\alpha = 0.05$, two-tailed) and determined that the difference in mean sentiment score between the two sets was indeed significant. This confirms our hypothesis that comments with the keywords "connect" and "battery" are more likely to have a negative sentiment. The results presented in Table 3 support this conclusion.

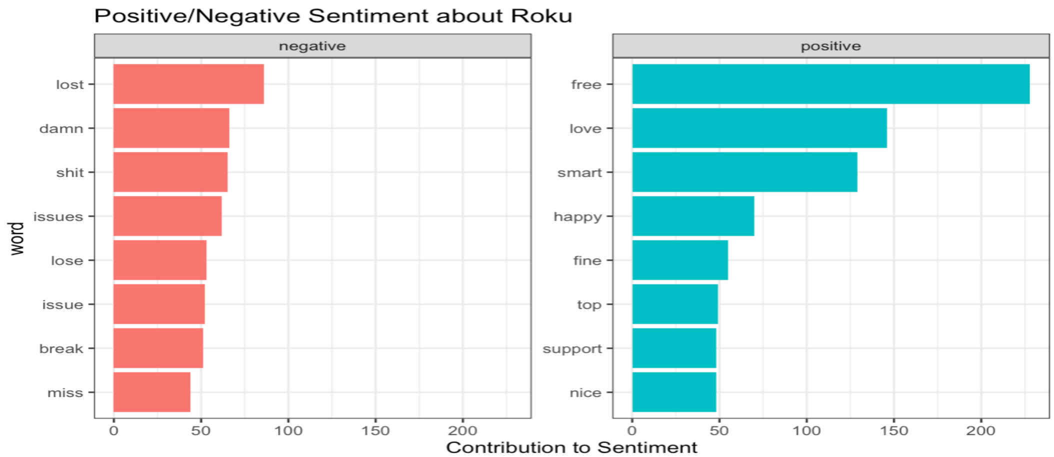
review with *“connect”*	**review without *“connect”***	**review with *“battery”***	**review without *“battery”***
0.4210	0.4890	0.3581	0.4857

[Table 3: Sentiment scores for comments containing the keywords “connect” and “battery”]

3. *Word Cloud Visualization and Sentiment Visualization*

We generated six word clouds [fig 4 and 5] using Python to compare keywords from two data sources, Twitter and Amazon reviews for each version of Roku product. We utilized the matplotlib, pandas, and word cloud libraries in Python to display the frequency of each word in the datasets.

[Fig 4: Word Clouds using Amazon reviews (top) and Tweets (bottom) for version OS 9.4, 10.0, 10.5 respectively]



[Fig 6: Visualization of the sentiment analysis]

We found that keywords related to product functionality, such as "remote", "channel", and "device", appeared in all three versions of Roku. The word clouds of tweets differed from the Amazon reviews, as company or brand names (e.g. Apple, YouTube, Amazon, Google) appeared more noticeably in the word clouds of tweets. This difference can be attributed to the public nature of Twitter, its user-friendly interface, and the possible regulation of online reviews. While word clouds provide a comprehensive visual representation of the data, they may not be the most accurate tool for comparison due to slight differences in the text cleaning and stemming steps in the pre-processing of tweets and Amazon reviews.

We then performed sentiment analysis using the Bing lexicon in R. We combined tweets for all three Roku versions and calculated the sentiment scores by counting the unique words and joining them with the sentiment data from Bing. The Bing lexicon categorizes words into positive and negative categories in a binary manner. Our analysis, shown in Figure 6, found that there were more positive tweets than negative tweets in the dataset. For example, no words in the negative sentiment category appeared more than 100 times, while positively categorized words such as "free", "love", and "smart" appeared more than 100 times, supporting our hypothesis that Roku products are positively perceived by users.

Results

We compared the key words generated by LDA with the update summary posted on the Roku official website to assess the impact of software updates on the user experience. The comments were divided into three categories: "Common features with frequent updates", "System and Hardware Connection Upgrades", and "New features and Contents". The table below summarizes our findings from the data sources of Amazon, Twitter, and the Combined data.

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. 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 are frustrated that the remote doesn't work even after resetting the device and have to resort to using the app to control Roku.	The Roku TV allows users to access free live broadcast TV with the use of their home's antenna.	The Roku is often compared to Apple TV.
OS 10.0	Users report that the earbud option on the remote drains the battery quickly.	No tweets were found discussing the features or sentiments towards OS 10.0.	No tweets were found discussing the features or sentiments towards OS 10.0 on Amazon or Twitter.
OS 10.5	The Roku can be combined in various ways to best fit the user's home setup and speakers.	Improvements have been made to the Roku mobile app, and users will experience faster channel launches.	No tweets were found discussing the features or sentiments towards OS 10.5 on Amazon or Twitter.

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	Improved WiFi connection Improved sound experience with Virtual Surround setting	Consumer Bought it to watch the nanny Join the club to watch a show	Not Found
OS 10.5	Automatic WiFi network detection for smooth streaming New interactive features and full surround sound	Easy access to cable alternatives, including Hulu + Live TV, fubo TV, Philo, Sling, and Youtube TV	Not Found

According to our analysis of Amazon reviews, the key factors in Roku's success include easy access to various cable options and content, seamless WIFI switching for better streaming, and convenient voice control and sound system. However, some unresolved issues were also noted, such as problems with the remote control not functioning even after resetting the device and the short battery life of the remote earbuds option. Our analysis of Twitter comments revealed that people buy Roku for its exclusive content, such as TV shows, live events, and movies. Overall, people compared Roku favorably to other brands due to its quick access and voice control features.

Our findings align with our hypothesis that Roku's frequent software updates have contributed to improved user experience. Although some comments contained negative keywords such as "issues," "break," and "miss," our sentiment analysis showed a positive overall score, suggesting that software updates have indeed enhanced the user experience.

Business Applications

Our project provides a comprehensive understanding of customer feedback on two major platforms using various analytical tools. It offers valuable insights for the company in three ways.

The first application is sentiment analysis, which gives an overview of how customers perceive the product or company. The company can monitor the sentiment score and take action if it deviates significantly from the normal range, such as investigating the reasons behind it, such as a product malfunction or new market opportunities.

The second application is pre-stage social feedback analysis, which identifies frequently discussed topics through the use of models like LDA and sentiment score analysis. This allows product managers to understand the underlying issues more efficiently by focusing on relevant comments. Our analysis found that connectivity issues and battery issues are frequently discussed, which Roku's management should take into consideration.

The third application is targeted advertising, which leverages the unique patterns identified on each platform. Our findings suggest that Amazon buyers tend to be more concerned with product quality, while Twitter users tend to compare Roku with major competitors such as Fire TV and Apple TV. Companies can use this information to create platform-specific content and conduct further research to determine the most effective content type for each platform.

Limitations

In this study, we faced several limitations in our data and methodology. Firstly, the datasets were collected within a specific timeframe, which may not fully reflect the current customer feedback. To address this issue in future studies, it would be helpful to set consistent time bounds across data collections.

Additionally, the accuracy of the topic modeling using Latent Dirichlet Allocation (LDA) was limited by the subjectivity of setting the number of topics. While certain keywords emerged from the LDA model, we cannot be certain that they accurately represent negative or positive reviews. Sentiment analysis also has its limitations in detecting sarcasm and slang, which may result in misinterpreting the sentiment of a review. For example, a review that contains negative words but is meant to be positive due to the use of sarcasm will be classified as negative.

Conclusion

This project aimed to gain insights into customer feedback regarding the Roku Ultra by analyzing reviews and tweets from Amazon and Twitter. From September 2020 to February 2022, we analyzed the data using techniques such as LDA, sentiment analysis, and word cloud visualization. Our findings showed that popular topics related to streaming, free content, TV, and remote control were commonly mentioned in customer feedback. We also discovered that system and hardware updates, as well as new features, were frequently discussed issues. These insights provide valuable information for the company to address customer concerns and improve their business strategies. However, further analysis with more data and advanced NLP techniques could be done to enhance the results of this project.