

Sentiment Analysis of Scraped Comments from r/NintendoSwitch

Michael Chung Ng

September 2020

Contents

1	Executive Summary	2
2	Methods	3
2.1	Initial data exploration	3
2.2	Assumptions	3
3	Results	4
3.1	Model Selection	4
4	Conclusion	5

Chapter 1

Executive Summary

We are interested in the popularity surrounding the Nintendo Switch video game system to inform the decision on whether a video game production company should invest in development of a game on the platform. To achieve this, we scraped comments from the r/nintendoswitch subreddit internet forum which serves as a discussion board for the Nintendo Switch. We used the PRAW (Python Reddit API Wrapper) API tool to scrap the comments and performed sentiment analysis on scraped comments to get an overall social sentiment surrounding the Nintendo Switch video game console.

The data set `comments-raw.csv` contains the data we have scraped from top posts on r/nintendoswitch with PRAW. Each row represents one reply to the main post; we have scraped the reply, number of upvotes it received and time the post was made, we were able to scrape 101826 comments.

After scraping the data we needed to perform the analysis, we were able to perform sentiment analysis with VADER (Valence Aware Dictionary and sEntiment Reasoner) and Textblob which are both python libraries that are pretrained to handle the processing of textual data and feed out an interpretable sentiment analysis.

From the sentiment analysis performed on the data gathered through scraping, we found that overall consensus surrounding the Nintendo Switch video game system was positive.

Chapter 2

Methods

2.1 Initial data exploration

All data was collected with the PRAW and scraped from the r/NintendoSwitch subreddit on reddit.com

Our initial data has four columns: `Reply`, `Upvote`, `Time`, `Key`. The reply corresponds to an individual comment, the upvote column is the number of upvotes the comment received, the time column corresponds to when the comment was submitted, and the key column corresponds to the post. Our data is made up of comments from 2017 - 2020; of which the majority of comments are from 2019 and 2020.

2.2 Assumptions

We are only scraping data from the Nintendo Switch subreddit which limits our data of getting a true unbiased sentiment of the product. We also have to understand that by scraping only from the Nintendo Switch subreddit we are creating an unwanted bias where we assume that people that visit this subreddit are generally fans and you would not expect a user critical of the product to visit a fan subreddit and instead voice their opinion on either a review website or elsewhere.

If we want to resolve this intrinsic bias, we would have to expand our scraping to a wider spectrum or potentially scraping reviews where we would be able to get a better idea of opinions and emotional sentiment of the consumers that have purchased the product.

Chapter 3

Results

3.1 Model Selection

To perform sentiment analysis on our data set, we chose to use VADER and Textblob which are two pre-trained natural language processing tools that are capable of performing the sentiment analysis. The reason we selected a pre-trained model over building our own classifier is that these natural processing tools have been validated and have been trained on immensely larger data sets which would prove time consuming for us if we chose to train our own classifier model to predict sentiment. The issue also with trying to build our own model is that if we approach this problem from a supervised perspective we need to have classified training data to be able to train the model.

Before the replies could be processed, we removed any URL links in them which would be irrelevant to the processing tools and pertain no sentiment. After cleaning up the replies, we tokenized the replies where each comment was split into individual words. Once the replies were cleaned, we fed the replies into VADER and Textblob and got an interpretable sentiment output.

We found that there was a drop in positive sentiment for the Nintendo Switch around the second quarter which could be caused by a lull in news or events for Nintendo Switch or this could be caused due to a low number of observations for that period

Chapter 4

Conclusion

The aim of this report was to inform a video game production company on whether they should go ahead with the development of a game on the Nintendo Switch video game system. We wanted to perform sentiment analysis to get an idea of the emotional sentiment towards the Nintendo Switch.

Using VADER and Textblob to conduct natural language processing and analysing sentiment, we got an overall positive emotional sentiment from users commenting in the Nintendo Switch subreddit. This tells us that the users have are likely having a positive experience with using their Nintendo Switch therefore we can inform the video game production with the go ahead with the development of their game on the Nintendo Switch platform.