

Taylor Swift In The Media

Written by Daniel Dai, Hubert Duhamel, Steven Koniaev, Youssef Samaan

¹McGill University

daniel.p.dai@mail.mcgill.ca
hubert.duhamel@mail.mcgill.ca
steven.koniaev@mail.mcgill.ca
youssef.samaan@mail.mcgill.ca

Abstract

In this study, we analyze American Media's Coverage of Taylor Swift in the media. We do so by retrieving relevant news articles from News API(News API 2023) and Gnews(Abdullah 2023). Using the retrieved articles we created a Typology and conducting an open-coding of these documents for both sentiment analysis and what topics the coverage is focused on. We explore the distributions of the categories amongst news sources and determine the most characteristic words for each category. Our open-coding produced several results. The Typography for the different kinds of coverage decided on is Slice of Life, Achievement, Fan, Music, Gossip, Tour, Controversy. Some of the most characteristic words for each category are Travis, hot, usa, version, la, film, rio respectively. We concluded that there is very little negative coverage of Taylor Swift. That NBC news covers the most controversial Taylor Swift stories and that most positive sentiment comes from the Achievement, Slice of Life, and Tour categories.

Introduction

The intersection between media representation and public perception plays a pivotal role in shaping narratives around public figures. This dynamic is especially pronounced in the realms of politics and entertainment, where individuals can experience dramatic shifts in public standing based on media portrayal. Analyzing media coverage and its tone in relation to public figures can yield fascinating insights. Taylor Swift, the focus of this paper, epitomizes someone profoundly impacted by media narratives. Since her debut at age 14, and now at 33, Swift has experienced the hardships of being a cultural and pop icon. She has been both celebrated and criticized by the media, experiencing the full spectrum of fame. In recent years, following various controversies, Swift has not been as prominent culturally. However, with her recent world tour and the rise of TikTok and alternative media platforms, her fan base, the "Swifties," have reignited interest in her work, propelling her back into the limelight. This research project methodically examines Taylor Swift's portrayal in North American news outlets, aiming to objectively assess the nature and emphasis of this coverage.

We employed two news aggregation APIs, GNews (Google News API) and News API, to compile a data set of 500 articles from a diverse range of English-language

sources in North America. This strategy was deliberately chosen to reduce potential bias in the representation and portrayal of Taylor Swift by the media during the data collection phase.

To analyze the data set, we adopted an open coding methodology, which enabled us to develop a typology for categorizing the collected articles. Additionally, we evaluated the sentiment of each article by analyzing its tone. This methodical approach ensured a systematic and unbiased examination of the content, allowing an objective analysis of the media coverage.

Data

The validity of our study relies on generating a data set that accurately reflects Taylor Swift's portrayal in North American media over the recent years. We therefore set out to collect 500 recent articles from different sources. In order to achieve a reliable data set, things such as data collection, data curation, data refinement and keyword selection were all explored and reasoned for our final data set to be applicable to our research.

Data Collection

In our initial data collection process, only News API was used in order to get all the necessary data as it is a comprehensive and also very powerful API for news data collection. However, this service employs a subscription model where for free accounts, we cannot look further back than 30 days. As the scope of the project is small we did not believe that paying the subscription fee was justified. Further in the time limit of 30 days, we were not capable to collect the 500 data point threshold that was set. Instead, to supplement our data set and achieve a perspective that goes further back than 30 days, we employed the Google News API, which provided access to articles from the past seven months. Further, since we had to manually decide the news sources for News API, the introduction of the GNews API permits us to retrieve articles from a wider range of news outlet. Since an article's content is often linked with the news outlet it comes from, this wider range of news sources let's us limit any kind of bias that could exist if we were to only use articles from News API. This dual-source approach allowed for a more extensive and varied collection of media representations.

Data Curation

After aggregating articles from both APIs, we removed duplicates, using article URLs as unique identifiers. This step was crucial to avoid redundancy and ensure the data set's cleanliness.

Keyword Selection

Since the purpose of research is to focus on media coverage on Taylor Swift, it was crucial to examine articles that mirrored the public's perspective, ensuring they were unfiltered and unbiased to accurately represent the viewpoint of the general population. Based on this reasoning, the keyword we used to search for articles were simply "Taylor Swift" since as a baseline, this was the most general keyword that would be used for anyone to get any information on Taylor Swift which is exactly what we want.

Data Refinement

While APIs generally provide relevant results, we encountered articles of little to no relevance to our question, such as those that only briefly mentioned Taylor Swift or were tangentially related to her. For example, we found an article with the title "JESSE WATTERS: The chemistry between Biden and Xi was unmistakable"(Staff 2023) which had nothing to do with Taylor Swift. The reason why this article was returned by the API is that Taylor Swift was briefly mentioned in the article, leading News API to identify the keyword Taylor Swift and return the article. Other example of such erroneous data that was returned included articles where they talked about a Taylor or a Swift but not necessarily Taylor Swift. To refine the data set, we implemented a filter to exclude any articles that did not feature "Taylor Swift" prominently in the title. This reasoning seems logical to us since generally, given how the media industry works, the title is meant to draw the readers attention and make them click on the article. Therefore, if an article doesn't include "Taylor Swift" in its title, then it's very likely that the article doesn't focus on her. Finally, since both Google News API and News API responses to a call are different, we had to reformat their responses so that all the data was uniform.

Final Data set

The final data set consists of 500 articles that focus on Taylor Swift, each formatted to include only pertinent metadata: source, URL, title, description, published date, and originating API. This formatting allows us to annotate the data easier and includes any information we might need for further analysis.

Methods

This study's methodological approach was designed on a pipeline, made to be both systematic and adaptable, making it possible to quickly and efficiently revisit and revise any stage if any abnormalities or optimizations are necessary. Given the nature of data science projects, the pipeline ideology is very important since abnormalities in the data is a common occurrence.

Pipeline Design

The pipeline design of our project is illustrated through figure 1.

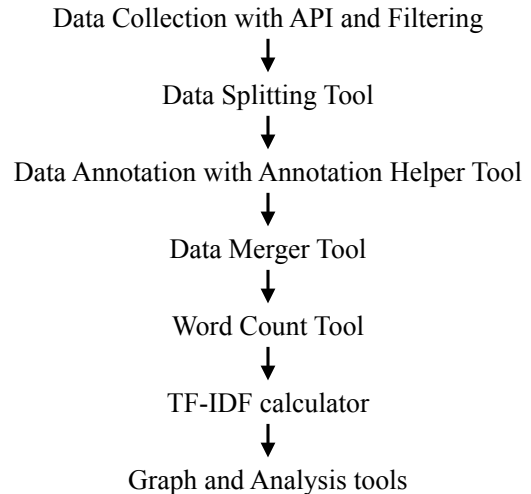


Figure 1: Research Pipeline for Media Coverage Analysis of Taylor Swift

Data Collection with API and Filtering

The initial data collection is made through a python script that collects the data from both News API and Google News API. The script then filters the articles returned by both APIs based on whether or not "Taylor Swift" appears in the title. It then formats the JSON returned by both APIs into a unified format and saves the data into the JSON file, `unified_swift_data.json`. More information about this step can be found in the Data section of the report.

Data Splitting Tool

The data set was then segmented into two parts: one designated for open coding and the other for subsequent annotation, following a 200/300 split. This segmentation was facilitated with the use of a tool we built that randomly splits articles into a specified number of separate JSON files. We introduced the randomness of the split since we did not want to risk introducing a bias in the coding where one of the annotator only had a few restrained type of categories compared to the others. Each of the 200 and 300 split were then again split into 4 splits. This second split was done since 4 annotators worked on the data and we wanted everyone to work equally. Further this method allowed for an iterative coding approach, which worked well to develop categories. More information about the categories found in this part of the pipeline can be found in the data category.

Data Annotation with Annotation Helper Tool

To streamline the open coding and annotation process, our team developed two specialized tools. One which facilitates category annotation and the other sentiment annotation. These tools enable efficient review and categorization of articles from designated JSON files, which are split and assigned to individual annotators. Each tool presents the annotator with the title and the description of the article, allowing for efficient annotation. The tool directly modifies the input JSON by adding a `category` and a `sentiment` field to every JSON element.

Data Merger Tool and Analysis Tools

Following annotation, each of the splits get aggregated into a single JSON file, `annotated_swift_data.json` using a data merger tool we developed. We then provided with the analysis of this data, where additional tools were implemented, including a word counter tool, which counted the amount of times certain words appeared in the title and description of articles for each of the categories in the typology, a TF-IDF calculator to determine the most characteristic words of each category, and a Jupyter notebook for statistical analysis and visualization of the data and its characteristics.

Version Control and Modularity

Although each component of the pipeline were connected to one another, they were designed to still be modular, allowing for quick and easy fixes to each component when we wanted to make some changes and optimizations to our approach. Further, git was used to coordinate the project's development and to facilitate cooperative development.

Results

Several key results have been found over the course of this study which demonstrates the positive view of Taylor Swift in the media and how different News Sources characterize her.

Open Coding

Our initial individual open codings had similar categories which were the following: 'Achievement', 'Music', 'Tour', 'Movie', 'Fan', 'N/A', 'Controversy', 'Personal-life', 'Fashion', 'Reputation', 'Public appearance'. Then, in the following coding, we had 'Achievement', 'Fan', 'Controversies', 'Public Appearances', 'Tour', 'Movie', 'Music', 'Personal-Life'. Subsequently, we found, 'Related', 'Celebrity', 'Music', 'Opinion', 'Quote', 'Fan' to better represent the data. Finally, our penultimate coding had the following categories: 'Gossip', 'Love Life', 'Accomplishment', 'Fan Related', 'Touring', 'Anecdotes (Jokes/stories)', 'Music', 'Personal Life'.

Typology

In the process of coming up with the final typology, we discussed the process we had done up until that point and the possible mistakes and edge cases that could occur within the previous choices of categories. After doing this exercise

and splitting and merging some categories, we came up with the final seven categories: 'Slice of Life', 'Achievement', 'Fan', 'Music', 'Tour', 'Controversy'. The final topics were defined as follows:

1. Slice of Life: Articles that have to do with Taylor Swift's Life (things like daily life, personal life, relationships)
2. Achievement: Articles that have to do with Taylor Swift's Achievements
3. Fan: Articles that have to do with the public's interactions with Taylor Swift as a brand and with other fans (including public companies)
4. Music: Articles that have about Taylor Swift's Music
5. Gossip: Articles that have gossip of information that exists from anyone relating to Taylor Swift
6. Tour: Articles about Taylor Swift's Tour (including the movie)
7. Controversy: Articles that have to do with controversies related to Taylor Swift

Characterization

Computing the TF-IDF score for each category over the documents we get the following most characterizing words.

1. Slice of Life: travis, kelce, jack, date, antonoff, plans, kelces, six, selena, photos
2. Achievement: no, 1, hot, hits, history, awards, artist, cruel, top, summer
3. Fan: usa, swiftie, ranked, jar, im, fans, special, hollywood, viral, yorker
4. Music: version', taylor's, '1989, version, taylors, rey, live, lana, hitting,del
5. Gossip: la, kelce's, harry, britney, biden, jason,'that's, wrote, singer, sheeran
6. Tour: film, tour', seattle, movie, brazil, concert, stage, eras, tickets, eras
7. Controversy: rio, fans, brazil, death, janeiro, heat, de, wave, 'i, springsteen

Exploratory Results

We plot the distribution of the topics over the dataset.

Slice of Life has the most coverage in the News while Controversy has the lowest. The second codings of the data gave us the distribution of Positive, Negative, and Neutral sentiments which were then plotted as well.

Both the Positive and Neutral sentiment articles have over 200 counts while the Negative articles are below 50. We then wanted to compare Sentiment with our topics to see which the correlation between sentiment and topics. Finally, we wanted to explore how different News Sources reported on Taylor Swift - if there are any outliers or certain coverage based on News Source. Many smaller News Sources were grouped into an Other category since they are irrelevant. Our findings show that the News Sources with the most controversial reporting of Taylor Swift is NBC News. While Fox News has the greatest coverage of the Slice of Life category. Fan and Gossip were heavily overed by PEOPLE and finally Music was covered geratly by Billboard.

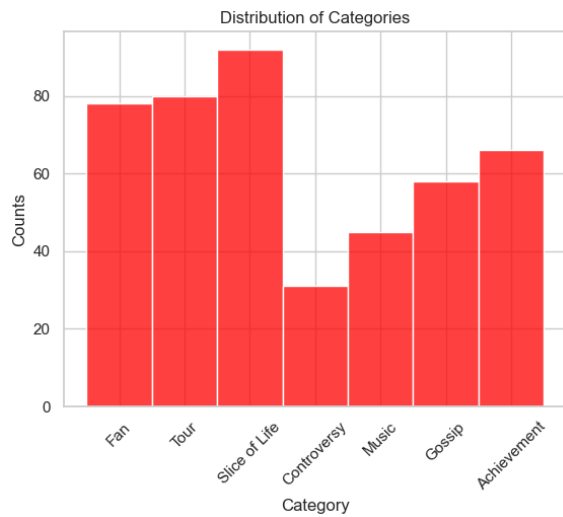


Figure 2: Distribution of the categories.

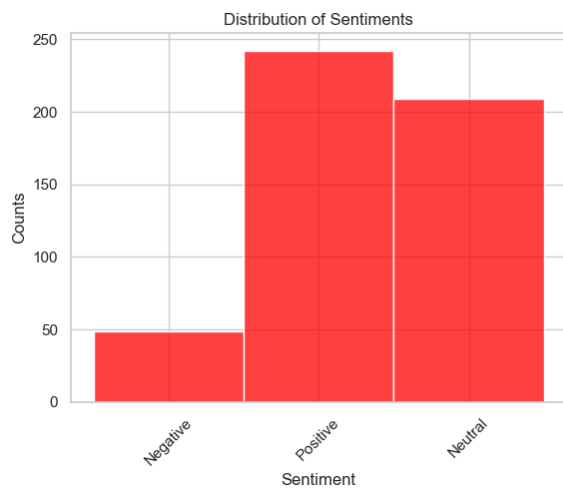


Figure 3: Distribution of the sentiments.

Discussion

Before discussing our results, it's important to outline the methodologies that could lend credibility to our analysis. Firstly, we employed multiple coding rounds to achieve a uniform understanding and application of coding criteria. This process was important since we wanted to remove any bias present in our individual minds and establish categories that were comprehensive and objective. This would let us all be on the same wavelength when it came to annotating our data.

The coding process began with open coding, during which we first individually determined the topics and coding strategies and then came together to discuss and come up with the final typology. This phase culminated in a final coding of 500 articles, the results of which are presented in the following section.

Additionally, in assessing the sentiment of the articles, we

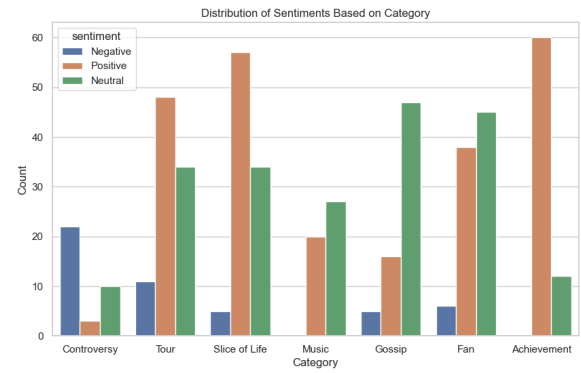


Figure 4: Distribution of the categories relative to the sentiment.

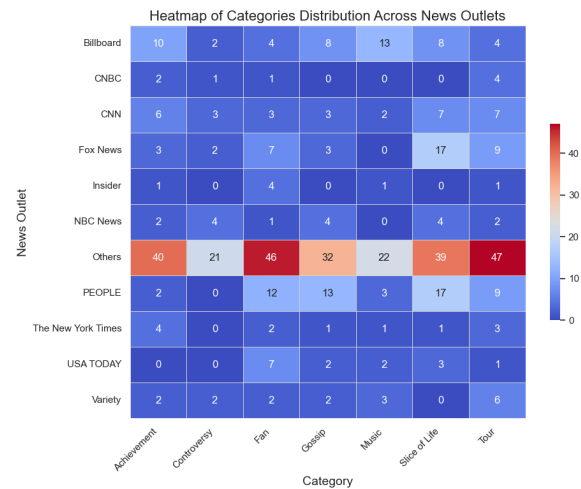


Figure 5: Heatmap of Topics relative to News outlets.

adopted the tone of said articles towards Taylor Swift. We analyzed whether the articles casted her in a positive light, portrayed her negatively or were neutral about her. Neutral articles were identified as those having no significant impact on her public reputation or image. It's noteworthy that many articles focused on Swift's recent 'Eras Tour' and her related movie/documentary.

Figure 2 Results

Figure 1 highlights a significant public interest in Taylor Swift's personal life, as evidenced by the volume of articles focused on her day-to-day activities. This suggests that such coverage is likely driven by reader engagement and profitability for news outlets. It is very well known that Taylor Swift's fans are very interested about her private life, more than general fans. They are know to dissect everything from the music she listens to the small interactions she has with other public figures, in order to create new theories about her upcoming albums, songs and promotions. It is therefore not a surprise that for media companies, the most profitable thing to do is to publish about Taylor Swift's private life. The next most frequent topics are her Eras Tour and fan-related

content. The abundance of tour articles comes from the fact that, at the time of writing this paper, her ongoing tour is breaking a number of records. Further her tour is generating an incredible amount of money, which is sure to peak the interest of the North American population since we are a capitalistic society. Notably, there is a higher prevalence of articles celebrating her achievements compared to those discussing controversies, although there is still a marked interest in gossip and her music career. At first glance, this distribution indicates that Taylor Swift is not currently receiving a lot of bad coverage from the press.

Figure 3 Results

Figure 3 reveals a predominantly positive portrayal of Taylor Swift in media outlets, with a substantial number of articles expressing neutral sentiments closely following. The scarcity of negative coverage is noteworthy. This trend could be attributed to a combination of factors: media outlets may be inclined to maintain a positive relationship with a high-profile celebrity like Swift, due to the possible backlash from Taylor Swift's fans towards either the media company or the author of the article. Although this is not much of a big problem for large media outlets, this could be a significant faux-pas for smaller media outlets which don't have the means to fully protect themselves in case of a big disagreement. It could also reflect her generally favorable public image and lack of actions that provoke negative responses from both the public and the media.

Figure 4 Results

The analysis of Figure 4 reveals that controversies involving Taylor Swift usually garner negative media attention, which is consistent with the typical perception of controversies. It is expected that controversies would cast individuals involved in a negative light. Conversely, Taylor Swift's accomplishments are often portrayed positively, aligning with the inherent positive connotation of the term 'achievement.' Interestingly, gossip articles about Swift often maintain a neutral tone, deviating from the common assumption that gossip is inherently negative. This unexpected neutrality could stem from the specific definition of 'gossip' used in our typology. Notably, Taylor Swift's music is exclusively associated with positive or neutral sentiments, underlining its favorable reception. Similarly, coverage of her tours and fans predominantly carries positive sentiments. Echoing earlier observations, it is crucial to recognize that media outlets are likely cautious about alienating Swift's dedicated fandom. While media companies wield significant influence, the potential backlash from Swift's fans is not to be underestimated, possibly explaining the scarcity of negative coverage about her.

Figure 5 Results

The heatmap analysis reveals a trend: nearly all major media outlets cover Taylor Swift's Tour. This widespread coverage is likely driven by the tour's status as a cultural phenomenon, capturing significant public interest. Consequently, news outlets are motivated to report on this topic, anticipating engagement and clicks. Additionally, entertainment outlets,

such as PEOPLE and Fox News, show a marked preference for covering 'slice of life,' 'gossip,' and 'fan'-related stories about Taylor Swift, more so than other media companies.

It's also noteworthy that a significant portion of Billboard's content categorized under 'achievements' are about their own awards, such as the Billboard Hot 100 and the Billboard Music Awards. Lastly, the heatmap indicates that news outlets generally produce a wide range of articles about Taylor Swift.

TF-IDF Score Analysis

The TF-IDF scores for each category in our analysis offer insightful revelations about the predominant themes in Taylor Swift's media coverage. In the 'Slice of Life' category, the prominence of words like 'travis' and 'kelce' highlights the media's and public's deep interest in Swift's personal life, particularly her relationship with Travis Kelce.

In the 'Achievement' category, terms like 'no' and '1' frequently appear, likely signifying her ranking in various music charts and awards. This aligns with the industry's metrics of success, where terms like 'top', 'award', and 'hits' are common indicators of achievements.

For the 'Fan' category, the term 'swiftie', the affectionate name for Swift's fandom, appears as a key word.

In the 'Music' category, the term 'version's' stands out, likely referencing the re-releases or new versions of her albums.

The 'Gossip' category is characterized by the names of other public figures, suggesting that much of the gossip in Swift's media coverage involves her interactions or relationships with other celebrities or what they think about her.

The 'Tour' category features words like 'movie', 'concert', and 'eras', pointing towards the extensive coverage of her tours and related projects, such as The Eras Tours Movie.

Lastly, the 'Controversy' category includes words like 'rio', 'death', and 'fans', alluding to specific incidents, such as the unfortunate event where a fan passed away at Taylor Swift's Era concert in Rio de Janeiro.

In conclusion, we see that using TF-IDF, we can both a high-level and generally precise view of what each category's content is about.

Conclusion

We analyzed American Media's Coverage of Taylor Swift in the media. Using North-American News APIs we were able to get 500 articles and conducted an open coding exercise on the data for both Sentiment and Topics. We found that the coverage of Taylor Swift is overwhelmingly positive and that most recent coverage focuses on her Personal Life and Tour. We were able to characterize the categories by looking at the highest TF-IDF scores for each topic. This allowed us to see which words are the most commonly used word to describe the category. We concluded that there is very little negative coverage of Taylor Swift. Overall, Taylor Swift has recently been prominent in the media and North American News outlets are in full support of her.

Group Member Contribution

Daniel's contributions: Daniel did 1/4 of annotations and helped with the open coding exercise, created annotation helper, annotation splitter and merger, and wrote the Data and Method sections of the report.

Hubert's contributions: Hubert did 1/4 of annotations and helped with the open coding exercise, coded the data collection and data unifier scripts, created figure 1,3,4 and 5, edited and proofread the report's content and structure, and wrote the final version of the Introduction and Discussion section of the report.

Steven's contributions: Steven did 1/4 of annotations and helped with the open coding exercise, implemented the Wordcount and TFIDF scripts, created figure 2, and wrote the Results, Abstract, Conclusion part of the report.

Youssef's contributions: Youssef did 1/4 of annotations and helped with the open coding exercise, created the sentiment coding helper, reviewed the report, and wrote the initial version of the Introduction and Discussion for the report.

References

Abdullah, M. 2023. GNews: Python Package for Google News Article Search. <https://github.com/ranahaani/GNews>. Accessed: 2023-11-19.

News API. 2023. NewsAPI.org. <https://newsapi.org>. Accessed: 2023-11-18.

Staff, F. N. 2023. Jesse Watters: The chemistry between Biden and Xi was unmistakable. *Fox News*. Accessed: 2023-11-22.