# How Much Sad Did You Think I Had in Me: An Evaluation of Taylor Swift's Musical Progression using Sentiment Analysis and Topic Modeling

**Abby Kotar**


## Abstract

Over the trajectory of her career, Taylor Swift has taken the world by storm. Using this Swift's lyrics to compute a sentiment analysis with TextBlob, VADER, and Flair as well as topic modeling using Latent Dirichlet Allocation, this project will explore how her musical themes changed in relation to her success as an artist. As a whole, Swift's music shifts away from romantic relationships to more complex themes.


## 1 Related Works

There has been a lot of research performing different sentiment analysis techniques on lyrics. Lyrics can sometimes be harder for a machine to interpret since they use deliberately nuanced and interpretative language (Choi et al., 2015). Kathleen Napier and Loir Shamir specifically looked at the Billboard Hot 100 and utilized a tone analyzer provided by IBM's Watson tool in order to sort the songs into different moods. The tone analyzer uses a Support Vector Machine (SVM) with a one-vs-rest method (Napier and Shamir, 2018). This allows for multiclass classification. This study was done on music from 1951 until 2016 and showed an overall mood shift to more negative tones, like anger, disgust, and fear. In this project, there will be a similar progression over time but with a much smaller sample size. Using only Taylor Swift songs could limit the scope of something like Napier and Shamir's study. The smaller sample size could cause less accurate results than what Napier and Shamir calculated. Also, IBM's tone analyzer tool is no longer available so I will be applying some newer techniques to achieve a similar result. Sentiment analysis mainly focused on good, bad, or neutral while now I am trying to dive a little deeper into Swift's lyrics than just good, bad, or neutral. While the Watson tone analyzer could have been a good start, it would not provide as detailed analysis as topic modeling.

Similarly, Enrico Laoh, Isti Surjandari, and Limisgy Ramadhina Febirautami performed an analysis on the top Spotify Indonesian songs (Laoh et al., 2018). These authors propose using latent Dirichlet allocation, which is considered the most popular topic modeling algorithm. This paper mentions using perplexity to examine the results, which is the inverse per-word likelihood (Laoh et al., 2018). This will likely be a good factor to consider in my own analysis. Perplexity could help us determine the number of categories, which will definitely be important for my own research. Kahyun Choi, Jin Ha Lee, Craig Willis, and J. Stephen Downie evaluate their LDA with Point wise Mutual Information (PMI), which has been found to be highly correlated with human understanding, unlike perplexity (Choi et al., 2015). They use Normalized PMI, which is between 0, meaning the words never occur together, and 1, which means the words always occur together (Choi et al., 2015).

Using popular lyrics, Napir and Shamir looked at the progression of popular music over time. Similarly, North, Krause, and Ritchie examined popular British music to see if topics within the songs seemed to make them more famous (North et al., 2021). They seemed to find that music and lyrics equally contributed to the popularity of a song (North et al., 2021). Music involved different scores about dance-ability, beats per minute, and other measurements taken from the actual music itself. In this research study, only the lyrics will be examined, hoping that typically the music is related to the lyrics in terms of genre and emotion. However, this could be a limiting factor, and maybe in a further study examining the music itself would have been an interesting way to further understand why Swift's most popular songs became so influential. In addition, based on Pettijohn et al, there has been research to show that music interests changes during different economic circumstances. This research was specifically done with pop music, which can be related to most of Swift's discography (Pettijohn, 2012). More upbeat dance songs are seemingly more popular during periods of little economic stress, where slower, contemplative songs become popular in periods of economic distress (Pettijohn, 2012). This could be another factor to consider when understanding how Swift's songs became popular, as she shifted genres and vibes many times throughout her career.

Topic modeling is a statistical approach to text analysis and can be a way to understand large amounts of text (Devi and Saharia, 2020). Kento Watanabe and Masataka Goto also examined different techniques to understand lyrics like mood estimation and topic modeling. They call the study of lyrics, lyric information processing (LIP) as lyrics offer unique rhyme schemes and storytelling that plain words do not have (Watanabe and Goto, 2020). In Watanabe and Goto's paper they discuss many different ways lyrics can be analyzed in a natural language processing context but I want to specifically focus on topic modeling. They explain that topic modeling does not require training on valence and arousal levels, like mood estimation does (Watanabe and Goto, 2020). Maibam Debina Devi and Navanath Saharia mention latent Dirichlet allocation (LDA) which was able to recognize the emotion in songs with 73% accurary (Devi and Saharia, 2020). LDA aims to connect documents to topics using a bag-of-words as a feature (Devi and Saharia, 2020).In an earlier study that Goto further expanded on with two other researchers, Kosetsu Tsukuda and Keisuke Ishida, they created a lyric based explorer based on topic modeling that improved upon latent Dirichlet allocation (LDA) that takes into account artist's taste when creating topic modeling (Kosetsu Tsukuda and Goto, 2017). This is a really interesting tool that could act as a good comparison to the results of this study after completing topic modeling. This tool provides multiple different topics and shows the user the songs that attribute each emotion (Kosetsu Tsukuda and Goto, 2017). While this tool is interesting, this study will focus more on the change in topic over time to see if there is a shift that correlates with Swift's success. A similar strategy will be used but instead of sorting by group, sorting by album.

Overall, there has been plenty of research on sorting lyrics to understand the mood. While this initially was more focused on tone analyzing, it has shifted to topic modeling, which allows for more precision in creating categories. It is also important to remember the sociological effect from music and how different time periods can place value on different musical tones and genres.

## 2 Methods

Using the Taylor Swift discography from Kaggle, the data must first be cleaned (Clark, 2024). Because Swift is currently rerecording her discography, there are multiple versions of certain albums. These re-recordings will be sorted out. This is because this study is focusing on the progression of her music and fame, which although her re-recordings are popular, they are not made in the initial time period. There is also a few songs that Swift released not on one of her albums. These are either features, like for Ed Sheeran, or stand-alone songs for movies, like Carolina from Where the Crawdads Sing. Lastly, Swift's latest album, *The Tortured Poets Department*, had not been released when the data analysis was started. Therefore, it is not included in this analysis. After sorting out these songs, the lyrics must also be cleaned. Currently, they include who singing, like "Kendrick Lamar:" as a lyric. Cues for who signing where filtered out manually, since there are only a few songs with collaborators. In addition, there were instances of "/", "[]", or other punctuation that does not add to the meaning. These were all filtered out.

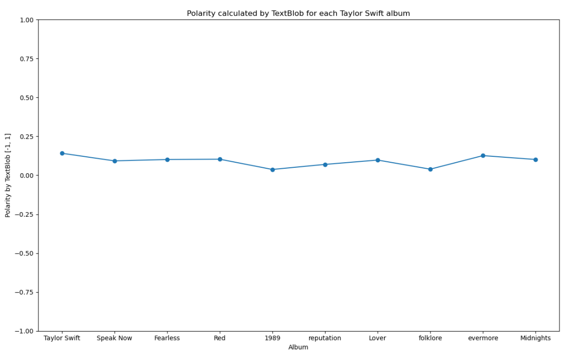
After data cleaning, there will be some simple sentiment analysis, somewhat similar to how Napir and Shamir looked at the progression of pop music over time (Napier and Shamir, 2018). There will be a few data visualizations to showcase this basic progression before diving into topic modeling and more complex analysis. This basic sentiment analysis will give a baseline for better understanding her lyric progression. There are many different Python libraries, like TextBlob, VADER, and Flair, that implement sentiment analysis. TextBlob is lexicon-based sentiment analysis which does not involve any machine learning. This assigns an orientation to each word and computes an average per sentence. VADER uses lexicon-based analysis in addition to a rule-based analysis. The rule-based analysis assigns tags to words. Flair offers an embedding based model, which might give a more accurate analysis of the lyrics. Using TextBlob, VADER, and Flair, the average sentiment per album will be calculated by first finding the polarity of each lyric. The polarity is defined on different scales for each library but its essentially a low score is very negative word association while a higher score is a greater word association. These average polarity scores are calculated per album and then graphed over the course of her ten studio albums. These models will be evaluated based on human-review of how accurate the sentiment seems.

After the simple analysis, this project will dive deeper into topic modeling. Like Laoh, Devi, and Saharia did, using latent Dirichlet allocation will be the best option. The Gensim library has a built in LDA model that will be used. First, to remove irrelevant words, Python's nltk library has a stopwords feature that contains a list of common stop words. Then, wordcloud was imported. After creating a wordcloud for each album, there was a vast number of words that did not contribute to the overall meaning of each album. Words like "know", "get", "I" were removed as well as sounds like "ha", "uh", "huh", and "mmm". After these were filtered out, additional word clouds were created to make sure most of the excessive words were removed. Then, LDA could occur on the lyrics in a bag-of-words representation of the text. The lyrics were put into a list of individual words. Using gensim.models.Phrases, the words could easily be converted into bigrams. Then, the words had to be lemmatized to group words with the same lemma together. Using code from Aravind CD's medium article, words from the lyrics were tagged with different parts of speech (CD, 2020). Since LDA lets the user choose the number of categories, there will be a few rounds with different numbers of categories. For each album, probabilities of being in each category will be examined to figure out how to model sorted her lyrics. Gensim's LDA takes in the lemmatized data set, a dictionary that has an id for each word, and the number of categories. These will be then evaluated by humans, as well as using gensim's log perplexity score and coherence score. Similar to Napier and Shamir, this analysis will be compared to her first week album sales (Napier and Shamir, 2018). Then Swift's progression could be compared to the Billboard's to see how societal music taste shifted.

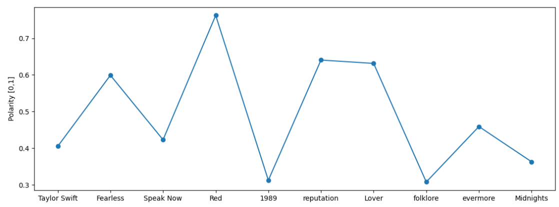
To evaluate sentiment analysis, it must be done by hand. Using my own knowledge of Swift and her lyrics, the validity of the models will be checked. For topic modeling with LDA, gensim provides two different ways to evaluate: perplexity and coherence. The log perplexity will be taken from the LDA model. Perplexity is how well a model predicts the sample. This could be minimized, as the model should be coherent and logical. Gensim also provides a coherence model, which takes in the LDA model, the corpus, and the type of coherence model to use. This project uses $c_v$ modeling, which is seen as the most human understandable way to evaluate topic modeling (Rijcken, 2023). In $c_v$ coherence, each topic word is compared to the set of all topics. It predicts the probability of two words co-existing in the same corpus (Rijcken, 2023). Unlike perplexity, coherence should be maximized. However, for both perplexity and coherence there is no defined threshold value for good results, so again I will be examining the topics myself to see what themes a human can pick out. These will be used to evalute how successful the topic modeling was to see if it can be used to truly show how Swift's lyrics have changed over the years.

## 3 Results

After data cleaning, the varying sentiment analysis were completed. First, TextBlob analysis was computed on each album. TextBlob computes a polarity between [-1, 1], in which -1 is the most negative emotion and 1 is the most positive emotion. Overall, TextBlob shows a pretty similar, generally positive shift in Swift's lyrics over the years. None of her albums dip into the "negative" polarity, although *1989* and *folklore* seem like to two saddest albums with Debut (*Taylor Swift*) or *evermore* being the happiest.

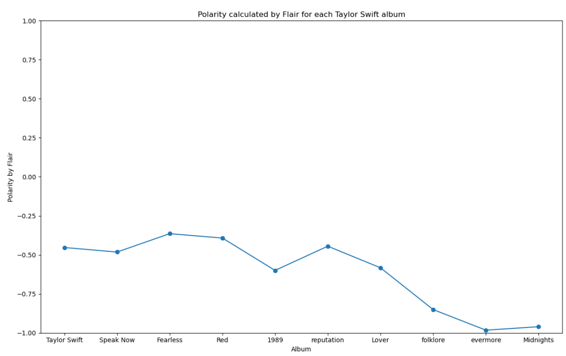



As mentioned earlier, TextBlob relies on a lexicon-based approach that does not take context into account. This is a reoccurring issue in the VADER analysis as well. VADER itself was trained on mostly social media data, which may not accurately translate to lyrics which are more poetic. VADER computes a polarity from [0,1], with 0 being the most negative emotion and 1 being the most positive emotion. This is a smaller range than TextBlob which makes the albums seem like they have more emotional difference. VADER similarly shows *1989* and *folklore* as the albums with the most negative emotions, while *Red* is the happiest album.




One issue both of these models have is not understanding context. For example, in Swift's *Shake it Off*, she says "Haters gonna hate, hate, hate, hate, hate." Both VADER and Textblob analyzed this as an incredibly negative lyric. Since it is repeated multiple times in this song, the *1989* score seems artificially low. From a human analysis, Swift's most depressing albums include *folklore*, *evermore*, and *Red*, which tend to be rated pretty high by both TextBlob and VADER. This could be because of her use of symbolism or lyricism that is just not understood by simpler models. To try and remedy this, Flair's more complex sentiment analysis was used.

Flair uses deep learning to better understand context and sarcasm. Flair uses a scale from [-1, 1], with -1 being the most negative emotion and 1 being the most positive emotion. Flair computes a positive emotional score and a negative emotional score. The positive and negative emotional scores were then averaged together for each album. This is seen below in Figure 3.




From this analysis, all of Swift's albums are ranked below 0, or mostly negative emotionally. Overall, Flair shows an increase in negative emotion over time, which seems to closer resemble the hypothesis that sad music is more popular over time. The album *evermore* and *Midnights* are particularly low, with evermore actually computing almost a -1. Still *1989* is shown as surprisingly low, when overall it is a seemingly happier album than *Red*. From these three libraries, Flair seems to be the most accurate in terms of its ability to comprehend context. Flair also captured the extreme sadness of *folklore*, *evermore*, and *Midnights* in particular, as well as the first shift in happiness after her initial growth from her debut album.


(Ammarr)

Compared to first week sales, her initial growth seems to grow with more positive lyrics. While *1989* is consistently reported as a sadder album than *Red*, sonically it is more pop-based, which could be a confounding factor in its greater first week sales than *Red*. Seemingly, it does not seem like a conclusion based on polarity in relation to sales can be made, as while *folklore* and *evermore* have lower sales and are sad, *Midnights* also ranks very negative but has the greatest first week sales. In addition, first week sales might not be reflective of the overall success of the entire album. However, Swift's success has also grown drastically with social media, which might have contributed to *Midnights*' commercial success. Looking at topic modeling may generate a fuller story of Swift's lyrical progression.

For topic modeling, LDA lets the user choose the number of categories. After a few trial runs with between 2 and 7 categories, 4 categories seem to lead to the lowest perplexity scores and the best human comprehension. Earlier, I removed all filler words, but ultimately it was hard to determine what words were filler and what words were meaningful. This was up to my discretion for the most part, as words like ”got,” ”they”, ”mine”, ”would've,” and other contractions were all removed. Some words like ”ever,” ”mine,” ”back,” may not seem like they have a lot of meaning but in a song, these words could represent different longing or wanting. Below are the topics created for each album with the most influential words listed.

```
Taylor Swift
Category  0 :
0.023*"burn" + 0.020*"let" + 0.019*"baby" + 0.018*"really" + 0.017*"time" + 0.016*"back" + 0.015*"song" + 0.014*"love" + 0.014*"drive" + 0.014*"bad"

Category  1 :
0.028*"think" + 0.019*"beautiful" + 0.015*"burn" + 0.015*"girl" + 0.015*"way" + 0.015*"love" + 0.013*"time" + 0.013*"old" + 0.011*"want" + 0.011*"let"

Category  2 :
0.042*"think" + 0.031*"back" + 0.027*"hope" + 0.020*"beautiful" + 0.013*"little" + 0.012*"find" + 0.012*"take" + 0.011*"time" + 0.011*"blue" + 0.011*"baby"

Category  3 :
0.027*"ever" + 0.026*"outside" + 0.022*"still" + 0.020*"break" + 0.019*"perfectly" + 0.017*"take" + 0.016*"back" + 0.016*"heart" + 0.014*"first" + 0.013*"good"
```

```
Fearless
Category  0 :
0.055*"come" + 0.023*"help" + 0.020*"night" + 0.018*"feel" + 0.015*"well" + 0.015*"one" + 0.015*"tell" + 0.015*"little" + 0.014*"name" + 0.012*"back"

Category  1 :
0.020*"way" + 0.018*"feel" + 0.017*"love" + 0.017*"baby" + 0.015*"fall" + 0.013*"belong" + 0.012*"time" + 0.011*"fearless" + 0.011*"take" + 0.011*"first"

Category  2 :
0.018*"back" + 0.017*"leave" + 0.016*"need" + 0.015*"baby" + 0.014*"time" + 0.014*"love" + 0.011*"door" + 0.011*"feel" + 0.010*"well" + 0.010*"window"

Category  3 :
0.016*"sorry" + 0.015*"time" + 0.015*"one" + 0.012*"tell" + 0.012*"take" + 0.012*"belong" + 0.012*"anymore" + 0.011*"love" + 0.009*"town" + 0.009*"baby"
```

```
Speak Now
Category  0 :
0.043*"mean" + 0.028*"long" + 0.024*"live" + 0.017*"big" + 0.017*"ever" + 0.017*"someday" + 0.012*"time" + 0.010*"back" + 0.010*"life" + 0.009*"hit"

Category  1 :
0.023*"back" + 0.023*"come" + 0.022*"long" + 0.021*"live" + 0.019*"time" + 0.016*"grow" + 0.013*"still" + 0.012*"day" + 0.012*"around" + 0.012*"ever"

Category  2 :
0.053*"back" + 0.036*"come" + 0.018*"think" + 0.015*"time" + 0.015*"mind" + 0.013*"still" + 0.011*"ever" + 0.009*"make" + 0.009*"nothing" + 0.009*"keep"

Category  3 :
0.020*"love" + 0.017*"away" + 0.016*"back" + 0.016*"meet" + 0.015*"ever" + 0.013*"grow" + 0.011*"forever" + 0.010*"run" + 0.010*"name" + 0.010*"around"
```

```
Red
Category  0 :
0.031*"starlight" + 0.027*"back" + 0.022*"time" + 0.020*"come" + 0.020*"stay" + 0.020*"love" + 0.019*"dancing" + 0.017*"look" + 0.017*"beautiful" + 0.015*"night"

Category  1 :
0.078*"red" + 0.035*"ever" + 0.033*"time" + 0.030*"stay" + 0.024*"back" + 0.020*"talk" + 0.018*"think" + 0.016*"together" + 0.016*"mad" + 0.013*"last"

Category  2 :
0.033*"time" + 0.032*"one" + 0.022*"tell" + 0.020*"last" + 0.017*"girl" + 0.016*"look" + 0.016*"home" + 0.013*"name" + 0.012*"right" + 0.011*"dance"

Category  3 :
0.053*"trouble" + 0.027*"time" + 0.021*"last" + 0.021*"ever" + 0.015*"love" + 0.013*"back" + 0.013*"right" + 0.013*"follow" + 0.011*"home" + 0.011*"put"
```

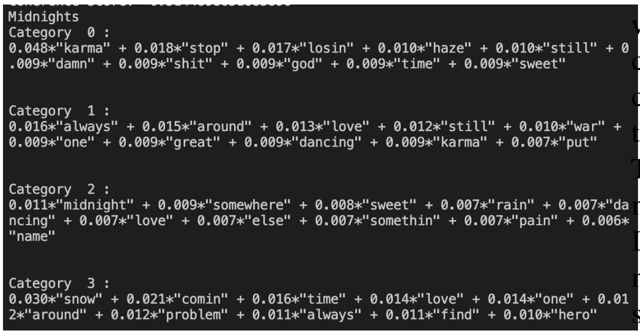
```
1989
Category  0 :
0.061*"love" + 0.042*"wonderland" + 0.026*"lost" + 0.026*"girl" + 0.024*"home" + 0.023*"way" + 0.021*"forever" + 0.019*"back" + 0.018*"silence" + 0.016*"ever"

Category  1 :
0.131*"wish" + 0.058*"stay" + 0.035*"shake" + 0.030*"back" + 0.023*"right" + 0.015*"want" + 0.013*"love" + 0.012*"come" + 0.011*"people" + 0.011*"hate"

Category  2 :
0.100*"yet" + 0.091*"shake" + 0.048*"clear" + 0.024*"fake" + 0.021*"good" + 0.021*"hate" + 0.020*"break" + 0.019*"play" + 0.017*"remember" + 0.015*"back"

Category  3 :
0.060*"love" + 0.040*"baby" + 0.036*"new" + 0.033*"bad" + 0.031*"blood" + 0.021*"think" + 0.017*"take" + 0.016*"look" + 0.013*"back" + 0.010*"come"
```

```
reputation
Category  0 :
0.033*"first" + 0.025*"time" + 0.024*"end" + 0.024*"call" + 0.023*"car" + 0.023*"game" + 0.020*"big" + 0.020*"baby" + 0.015*"love" + 0.015*"want"

Category  1 :
0.028*"waitin" + 0.026*"baby" + 0.021*"hold" + 0.017*"soul" + 0.016*"heart" + 0.016*"one" + 0.016*"body" + 0.015*"let" + 0.012*"nice" + 0.012*"time"

Category  2 :
0.033*"take" + 0.027*"good" + 0.022*"want" + 0.021*"bad" + 0.021*"light" + 0.019*"dress" + 0.016*"time" + 0.014*"feel" + 0.013*"somethin" + 0.013*"name"

Category  3 :
0.061*"look" + 0.038*"call" + 0.037*"want" + 0.022*"gorgeous" + 0.015*"feel" + 0.015*"face" + 0.015*"good" + 0.014*"bad" + 0.012*"back" + 0.012*"make"
```

```
Lover
Category  0 :
0.020*"street" + 0.020*"want" + 0.017*"one" + 0.016*"baby" + 0.016*"love" + 0.013*"walk" + 0.013*"forgot" + 0.010*"hate" + 0.008*"night" + 0.008*"need"

Category  1 :
0.031*"want" + 0.020*"think" + 0.020*"still" + 0.020*"love" + 0.018*"bless" + 0.015*"worship" + 0.013*"street" + 0.013*"take" + 0.013*"baby" + 0.013*"one"

Category  2 :
0.049*"man" + 0.023*"daylight" + 0.018*"baby" + 0.017*"one" + 0.014*"soon" + 0.012*"sick" + 0.011*"love" + 0.011*"want" + 0.011*"eeh" + 0.010*"comin"

Category  3 :
0.031*"love" + 0.030*"daylight" + 0.022*"one" + 0.021*"need" + 0.018*"right" + 0.018*"want" + 0.011*"fancy" + 0.011*"boy" + 0.011*"darling" + 0.008*"stay"
```

```
folklore
Category  0 :
0.028*"time" + 0.012*"hope" + 0.012*"around" + 0.012*"think" + 0.012*"call" + 0.010*"sign" + 0.010*"mine" + 0.010*"many" + 0.010*"back" + 0.009*"film"

Category  1 :
0.020*"time" + 0.013*"one" + 0.011*"back" + 0.009*"thing" + 0.009*"party" + 0.008*"pulled" + 0.008*"mine" + 0.007*"hope" + 0.007*"think" + 0.007*"mad"

Category  2 :
0.017*"time" + 0.014*"still" + 0.014*"think" + 0.012*"love" + 0.012*"give" + 0.011*"one" + 0.009*"want" + 0.009*"around" + 0.009*"come" + 0.008*"enough"

Category  3 :
0.015*"watch" + 0.012*"mad" + 0.010*"one" + 0.010*"still" + 0.009*"think" + 0.009*"breathe" + 0.009*"time" + 0.009*"want" + 0.009*"around" + 0.008*"serve"
```

```
evermore
Category  0 :
0.020*"think" + 0.017*"left" + 0.016*"evermore" + 0.012*"man" + 0.011*"prove" + 0.010*"take" + 0.009*"day" + 0.008*"catchin" + 0.008*"hand" + 0.008*"right"

Category  1 :
0.017*"call" + 0.012*"house" + 0.012*"road" + 0.012*"real" + 0.012*"good" + 0.011*"even" + 0.010*"always" + 0.010*"write" + 0.010*"hometown" + 0.009*"season"

Category  2 :
0.024*"thing" + 0.018*"right" + 0.016*"sometimes" + 0.013*"think" + 0.013*"left" + 0.013*"stay" + 0.011*"time" + 0.010*"dorothea" + 0.009*"still" + 0.008*"soul"

Category  3 :
0.017*"one" + 0.014*"love" + 0.013*"right" + 0.013*"long" + 0.012*"happiness" + 0.011*"time" + 0.009*"story" + 0.009*"short" + 0.009*"stay" + 0.008*"leave"
```

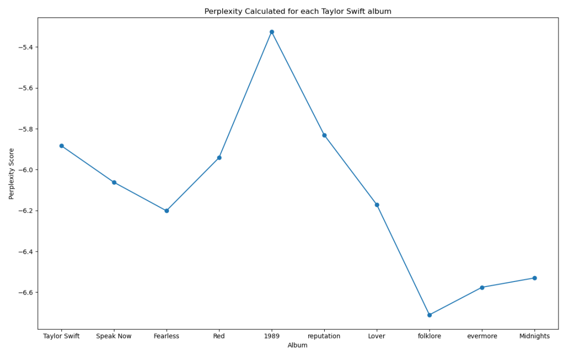


```
Midnights
Category  0 :
0.048*"karma" + 0.018*"stop" + 0.017*"losin" + 0.010*"haze" + 0.010*"still" + 0
.009*"damn" + 0.009*"shit" + 0.009*"god" + 0.009*"time" + 0.009*"sweet"


Category  1 :
0.016*"always" + 0.015*"around" + 0.013*"love" + 0.012*"still" + 0.010*"war" +
0.009*"one" + 0.009*"great" + 0.009*"dancing" + 0.009*"karma" + 0.007*"put"


Category  2 :
0.011*"midnight" + 0.009*"somewhere" + 0.008*"sweet" + 0.007*"rain" + 0.007*"da
ncing" + 0.007*"love" + 0.007*"else" + 0.007*"somethin" + 0.007*"pain" + 0.006*
"name"


Category  3 :
0.030*"snow" + 0.021*"comin" + 0.016*"time" + 0.014*"love" + 0.014*"one" + 0.01
2*"around" + 0.012*"problem" + 0.011*"always" + 0.011*"find" + 0.010*"hero"
```

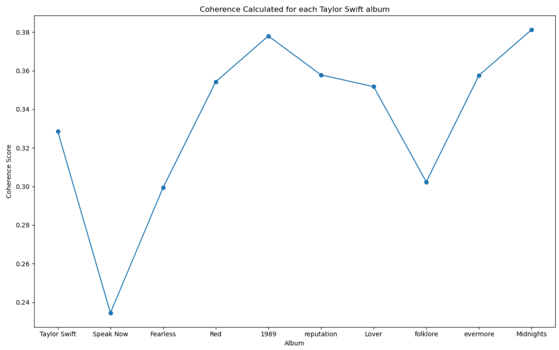
Over all albums, there is a continuous representation of love. While this transitions throughout the albums, ”love” is mentioned in at least one category for all albums. Focusing on how love is portrayed in each of her albums can show a potential thematic shift over time. *Taylor Swift*, *Fearless*, *Speak Now*, and *Red* can be grouped together for this transition, as they are her four original country albums. *1989*, *reputation*, and *Lover* are her initial shift to pop, with *folklore* and *evermore* being more alternative. *Midnights*, however, is a pop album again. Using these categories, her country albums relate ”love” to words like ”beautiful”, ”burn”, ”baby”, ”fearless”, and ”grow.” *Red* has two love categories, one focusing on more positive sounding words like ”starlight,” ”stay,” ”dancing,” and ”beautiful” while the other includes ”trouble,” ”last”, ”back,” and ”follow.” While previous albums have break-up songs that can be categorized as sad or mad, *Red* contains sad love in a more longing context. As her pop career starts with *1989*, ”love” is again in two categories: one with ”wonderland,” ”lost,” ”home,” ”forever,” and ”silence” and the other with ”baby,” ”new,” and ”bad.” These seem thematically closer in terms of being lost in love and brand new love. *Lover*, which focuses mostly on love, has love in all four categories with seemingly different contexts. However, *folklore*, *evermore*, and *Midnights* were deemed Swift’s albums with the most negative emotions by Flair. In these albums, love seems to be associated with ”stay,” ”time,” ”hero”, ”right,” ”one,” and ”time”, which seem to be somewhat consistent with the perceptions of love in her earlier music. The use of love throughout her discography shows a shift in how she portrays love and the emotions caused by it. Her most successful pop albums tend to have one category for exciting and new love, which is similar to *Fearless*. However, the difference could also come from Swift aging, as her country albums were younger and employ a more straight forward view of love, either break up or happy in love. In comparison, her newer albums tend to have more complex themes about love, either fighting in love, the idea of being lost, and more mystical elements. This less idealized idea of love could make her music more relatable to the general audience.

In Swift’s transition into pop, her music seems more repetitive and following a typical pop-structure with a repeated chorus. In her two best selling albums, *1989* and *Midnights*, there are song titles that appear in her the topic modeling such as ”bad” and ”blood” (*Bad Blood*) in *1989* and ”karma,” (*Karma*), ”haze” (*Lavender Haze*), and ”snow” (*Snow on the Beach*) from *Midnights*. This could show that a ”catchy” chorus that is repetitive could be more successful as these words become more influential in a category. In addition, *Midnights* is the only album with a swear word in any category. Both ”damn” and ”shit” appear in *Midnights* category 0. This could cause a lower polarity from the sentiment analysis. The category puts them with ”karma,” ”god,” ”losin,” and ”haze” which is a more mystical theme than her other albums.

In terms of commercial success and topics over time, Swift’s initial growth seemed to peak at *1989*, which had her greatest first week sales until *Midnights*. After analyzing each individual category, there are some categories that are not very human comprehensible. Ultimately, gensim built in log perplexity for LDA models was used to calculate perplexity. This should be minimized. As seen in the graph below, *folklore*, *evermore*, and *Midnights* seem to have the lowest perplexity.




Then, using gensim’s coherence model, coherence was calculated using the $c_v$ model. Coherence should be maximized. In this case, *1989* and *Midnights* have the highest coherence. *1989* seems to have one of the highest coherence scores and highest perplexities, which was unexpected.




When looking at the individual categories with a human eye, there are a few that are indistinct. They do not seem to have a theme that I am able to pick up on. However, using the categories that were distinct, there is a shift in Swift's music over time. Her initial two albums, *Taylor Swift* and *Fearless*, both have one or more categories focused on romantic break-ups or heartbreak. These can be divided into sad and mad break-up heartbreaks. In addition, there is also a category for nostalgia and regret. *Speak Now* shows a shift, where a majority of the categories are about growing up and nostalgia. This album sold more records than the previous two albums, and maybe this shift sparked a change. However, *Red* seems to go back to these previous themes of sad breakup and mad breakup, as well as love. Swift's peak at the time, *1989*, had themes categorized as love (more mystical than before with mentions of wonderland) and new love. The other categories seem more unclear, with words like "shake," "wish," "break," and "play." These seem more fun and upbeat than previous albums, however. *reputation* shows another shift in the love category to more soulful and sensual. Also, there are themes of partying and play again, similar to *1989*. *Lover* shows another dramatic shift, as all categories mention love in some capacity. They can be sorted into a complicated love, devoted love, and then two for seemingly more generic love. *1989*, *reputation*, and *Lover* are Swift's most pop based albums and tend to show more playful themes. *folklore* and *evermore* are Swift's shift to somewhat alternative or folk music where themes of nostalgia, sad love, home, and generic love are present again. Lastly, in Swift's return to pop, *Midnights*, has themes of karma/ mystics, complicated love, dancing, and generic love. Like *1989*, both of these albums have a more "mystical" category, which could be due to Swift's lyricism. Looking at the data from sentiment analysis as well, *1989* keeps coming up as the most negative album. While this may be due to lyrics aforementioned, it could also be that Swift uses a pop beat to "cover up" negative emotions. *Midnights* also ranked very negative with the Flair analysis, which could verify this assertion.

Overall, Swift seemed to shift away from breakup songs into topics associated more with happiness, like dancing. *Taylor Swift*, *Fearless*, and *Red* have much clearer "break up" associations, which seems like Swift also transitioned away from singing about romantic separation. This could be a response to her initially being known as only singing about her ex-boyfriends. In her later albums, there is more diverse themes, except for *Lover*. This thematically makes sense based on the title of the album, as Lover is mostly about different types of love. However, love in general, although the types of love and ways she sings about it has changed, still remains a constant theme in Swift's music. As Swift rises in popularity, it is clear that her themes have become more complex and varied. This could be due to catering to a wider audience as her fans or her own growth as a person and artist.

One limitation that affected the scope of this research is the data from album sales. Both *folklore* and *evermore* were surprise albums, with basically no advertising or build up to their release. Especially for *evermore*, this was seen in its lack of sales. Also, *Midnights* had multiple album variants that could be pre-ordered online. Fans who wanted to collect all of them could have bought multiple each, which resulted in a higher number of sales. The sales data also does not contain streaming, which has become an increasing popular way to listen to music. The sales on the graph are also only reported in the United States, while Swift does have a large international fan base. In addition, this project does not contain any analysis of *The Tortured Poets Department*, which has outsold all of Swift's previous albums. Her growth also was impacted by things besides her lyrics, like the famous Kanye West feud, the pandemic, the use of social media and the rise of Tik Tok. All of these factors could have added to or potentially decreased her album sales. Also, her re-recordings could have contributed to her rise in fame. Her re-recording process was greatly embraced by fans and could draw fans in who previously only enjoyed her old work. While her lyricism has clearly evolved, the hypothesis that sad or negative lyrics made her more popular is not clearly seen by these results. Instead, the analysis shows a more complex growth in her tones relating to love and life in general. Rather than a negative or positive binary, Swift uses all emotions to relate to millions of fans.

## References


Ammarr. [link].

Aravind CD. 2020. Topic modeling using gensim-lda in python.

Kahyun Choi, Jin Ha Lee, Craig Willis, and J. Stephen Downie. 2015. Topic modeling users' interpretations of songs to inform subject access in music digital libraries. In *Proceedings of the 15th ACM/IEEE-CS Joint Conference on Digital Libraries*, JCDL '15, page 183–186, New York, NY, USA. Association for Computing Machinery.

Madeline Clark. 2024. Taylor swift - released song discography (genius).

Maibam Debina Devi and Navanath Saharia. 2020. Exploiting topic modelling to classify sentiment from lyrics. In *Machine Learning, Image Processing, Network Security and Data Sciences*, pages 411–423, Singapore. Springer Singapore.

Keisuke Ishida Kosetsu Tsukuda and Masataka Goto. 2017. Alternation. *In Proceedings of the 18th International Society for Music Information Retrieval Conference*, pages 544–551.

Enrico Laoh, Isti Surjandari, and Limisgy Ramadhina Febirautami. 2018. Indonesians' song lyrics topic modelling using latent dirichlet allocation. In *2018 5th International Conference on Information Science and Control Engineering (ICISCE)*, pages 270–274.

Kathleen Napier and Lior Shamir. 2018. Quantitative sentiment analysis of lyrics in popular music. *Journal of Popular Music Studies*.

A. North, Amanda Krause, and David Ritchie. 2021. The relationship between pop music and lyrics: A computerized content analysis of the united kingdom's weekly top five singles, 1999–2013. *Psychology of Music*, 49:735–758.

Eastman J.T. Richard K.G. Pettijohn, T.F. 2012. And the beat goes on: Popular billboard song beats per minute and key signatures vary with social and economic conditions. *Curr Psychol 31*, pages 313–317.

Emil Rijcken. 2023. C topic coherence explained.

Kento Watanabe and Masataka Goto. 2020. Lyrics information processing: Analysis, generation, and applications. *Association for Computational Linguistics*.