# Conducting a sentiment analysis on Taylor Swift to understand public opinion and perception

Mikayla Peterson

# Contents

# How has public perception of Taylor Swift changed since the announcement and opening shows of the *Eras Tour*?

# Hypothesis

I think that for the brief period before the official announcement of the Eras Tour on November 1st, 2022, the overall sentiment is positive. The Ticketmaster Verified Fan Presale for the Eras Tour began November 15th, 2022; due to major complications with the presale, I think public sentiment shifted negatively.

The opening shows of the tour were in Glendale, AZ on March 17th and 18th and the tour has received glowing reviews. As a result, I think public sentiment became more positive.

I will be using sentiment analysis to test my hypothesis.

ENTERTAINMENT > MUSIC

## Taylor Swift's Eras Tour: A Timeline of the Ticketmaster Fiasco

From the presale chaos to the response from Swifties and lawmakers, here's what to know about the debacle surrounding Ticketmaster and Taylor Swift's latest tour

By Karli Bendlin | Updated on January 25, 2023 09:26 AM

RollingStone MUSIC POLITICS TV & MOVIES (SUB)CULTURE RS RECOMMENDS

LIVE REVIEW

## Taylor Swift's The Eras Tour Is a 3-Hour Career-Spanning Victory Lap

"So tonight, we're going to be going through an adventure, one era at a time," Swift tells Glendale audience ahead of odyssey into her catalog

BY WAISS DAVID ARAMESH

MARCH 18, 2023

# Sentiment Analysis

# Data Collection

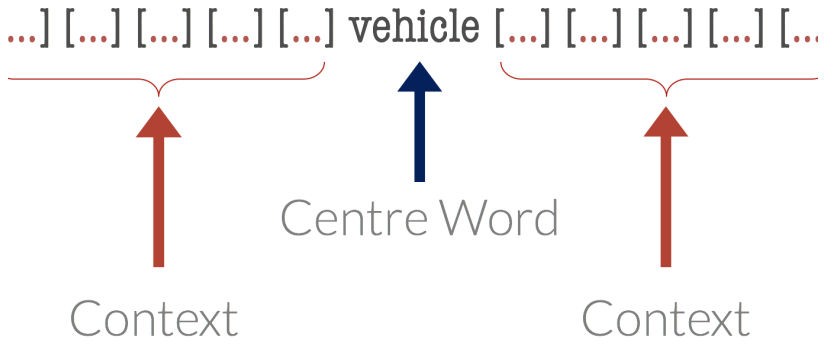
- Timespan: October 21, 2022 – March 21, 2023
- 2 categories: tweets ‘to’ Taylor and tweets mentioning ‘@taylorswift13’
- Used a scraper called SNscrape
- 600,000 tweets [a small portion, ~1000 tweets, will be classified manually to train and classify using Naïve Bayes, and the rest will be classified with Vader since I’m worried about finishing the project on time]

| | ID | URL | Date | User | Tweet | Retweets | Likes | Replies | Quotes |
|---|---|---|---|---|---|---|---|---|---|
| **149999** | 149999 | https://twitter.com/stainlessswift/status/1583... | 2022-10-21 | stainlessswift | @taylorswift13 tell me when something is | NaN | NaN | NaN | NaN |
| **149998** | 149998 | https://twitter.com/ParisCole9 | | | | | | | |
| **149996** | 149996 | https://twitter.com/Lou_Mc_Corr | | | | | | | |
| **149997** | 149997 | https://twitter.com/justinxbff | | | | | | | |
| **149995** | 149995 | https://twitter.com/greyymae/st | | | | | | | |

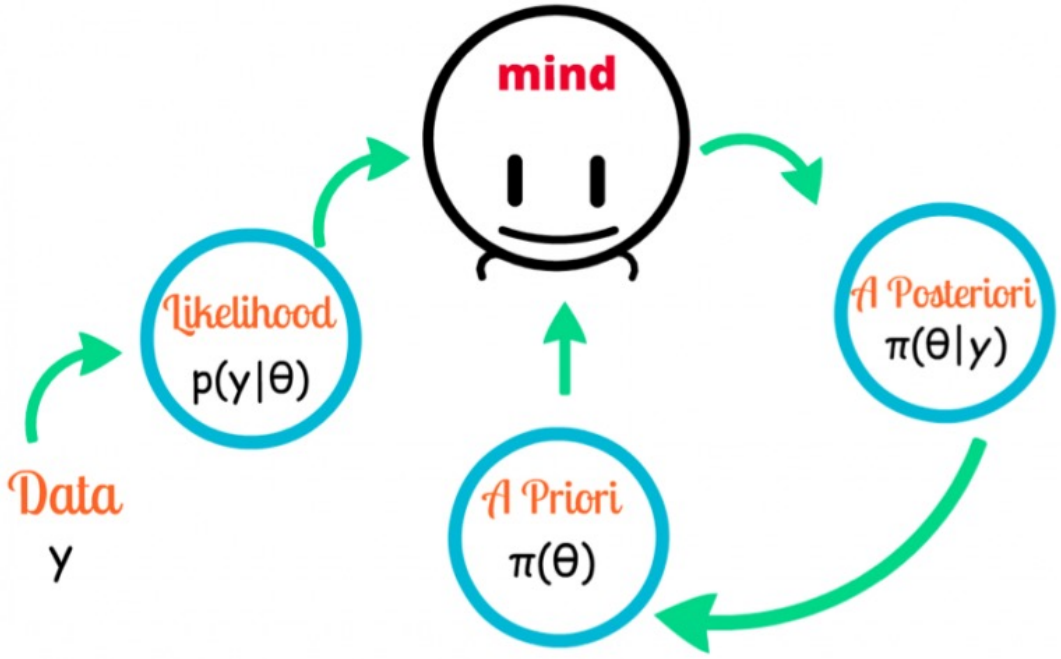
| | ID | URL | Date | User | Tweet | Retweets | Likes | Replies | Quotes |
|---|---|---|---|---|---|---|---|---|---|
| **450004** | 4 | https://twitter.com/sweetnegrande/status/16380... | 2023-03-21 04:39:46+00:00 | sweetnegrande | @airhe4dd_ @corneil_ri @PopCrave @taylorswift1... | 0.0 | 0.0 | 0.0 | 0.0 |
| **450003** | 3 | https://twitter.com/robonthemic/status/1638037... | 2023-03-21 04:40:08+00:00 | robonthemic | @USCmensvolley @uscwomensvolley @taylorswift13... | 0.0 | 0.0 | 0.0 | 0.0 |
| **450002** | 2 | https://twitter.com/reachonair/status/16380378... | 2023-03-21 04:40:33+00:00 | reachonair | Come and listen to Message In A Bottle (Taylor... | 0.0 | 0.0 | 0.0 | 0.0 |
| **450001** | 1 | https://twitter.com/starsglow_/status/16380379... | 2023-03-21 04:41:01+00:00 | starsglow_ | @getawaydynasty @WerePagesTurned @TheSwiftSoci... | 0.0 | 0.0 | 0.0 | 0.0 |
| **450000** | 0 | https://twitter.com/Insidelcarus/status/163803... | 2023-03-21 04:41:07+00:00 | Insidelcarus | @tis_sjoberg @shekharmehndole @taylorswift13 I... | 0.0 | 0.0 | 0.0 | 0.0 |

# Preprocessing*

- Natural Language Toolkit
- Each tweet is tokenized, words are replaced with their base form, ‘stop words’ are removed, and overall noise i.e. hyperlinks must be dealt with
- My dataset will be split into training and test sets with an (80/20)




How do I install a hard disk drive?
How do I use Adobe Photoshop?
Where can I learn more about computers?
How to download a video from YouTube
What is a special character?
How do I clear my Internet browser history?
How do you split the screen in Windows?
How do I remove the keys on a keyboard?
How do I install a hard disk drive?




# Machine Learning Algorithm

- Vader lexicon-analysis for most of the dataset
- Using Naïve Bayes to classify tweets as positive, negative, or neutral
- Naïve Bayes is good for classifying large amounts of data, but it is bad for predicting future values due to how it estimates probabilities
- I am currently repurposing the Naïve Bayes implementation I had to do for a lab in ENSE 412, Machine Learning
- I will compare it to the built in one in the Natural Language Toolkit to ensure my model is accurate; if my model is not accurate, I will use the built in one

# Naïve Bayes Crash Course

- Multi-classifier generative classification approach: the model of the distribution of input characteristics of the class is made using Bayes' classifier
  - Discriminative classifiers estimate decision boundaries from labeled examples
- We have to assume the features are independent given the class
- Use the decision rule to figure out which class the datum belongs to

Likelihood

Class Prior Probability

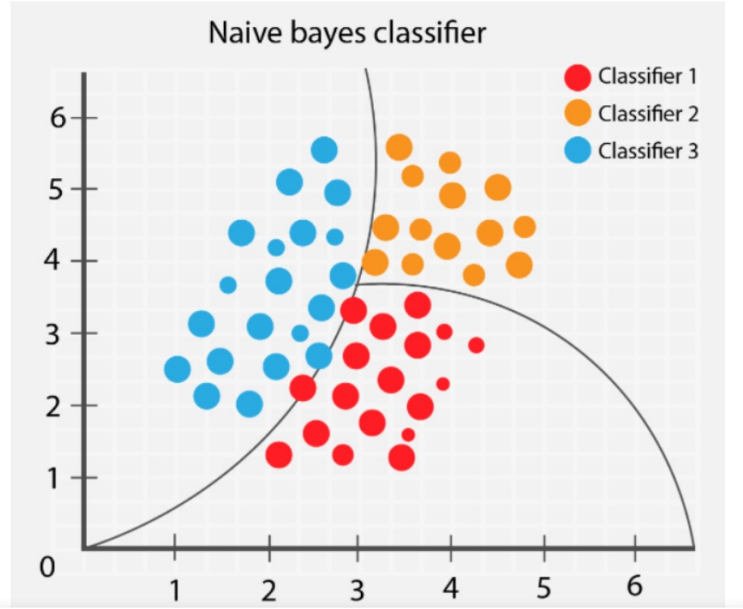
$$P(c \mid x) = \frac{P(x \mid c)P(c)}{P(x)}$$

Posterior Probability

Predictor Prior Probability

$$P(c \mid \mathrm{X}) = P(x_1 \mid c) \times P(x_2 \mid c) \times \cdots \times P(x_n \mid c) \times P(c)$$

$$y = \arg\max_k p(\mathbf{x}|t = k) = \arg\max_k \prod_{i=1}^{d} p(x_i|t = k)$$

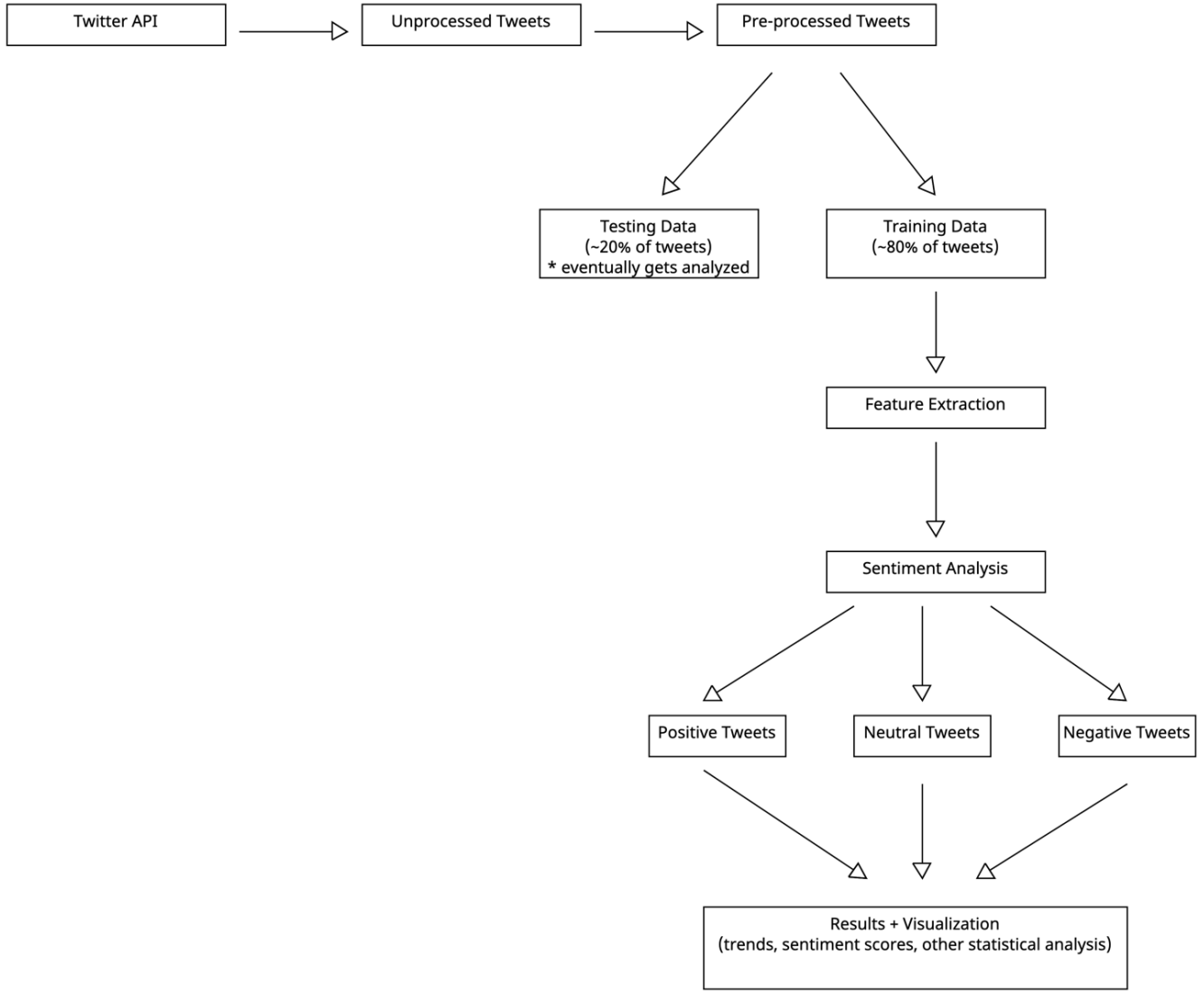


# Visualization/Samples

- Throughout my notebook, I will be putting various graphs to visualize the data
- I am currently working on an interactive dashboard with Google Looker Studio
- Since these are a work in progress, I can't show these yet
- I also don't have any samples to show quite yet since some of my data is still being preprocessed

# Structure Diagram

- Files involved: Jupyter Notebook and the datasets in CSV files



# How to Use

- All the code is contained in a single Jupyter Notebook
- All the other libraries I'm using are listed at the top of the Notebook along with installation instructions for the libraries
- In order to run the project, you can download new datasets manually (not recommended) or you can make sure that my predownloaded datasets are in the same directory as the notebook and then it will run
- The dashboard link is listed in the Notebook and will be made public when it's done

# Conclusion (So Far)

- I have not proven/disproven my hypothesis yet since I'm not quite done the system yet.
- I did not accomplish anything new with respect to Machine Learning/Lexicon based analysis
- Pro:
  - The approach I'm using is a very common way to do sentiment analysis so there are a lot of tutorials on how to do it
- Cons:
  - Since 'noise' in the tweets has to be tailored to each project, I feel like I could have done a better job trying to filter out the noise
  - I'm not sure how well the classifier handles slang so this should be looked into in the future
  - Gathering the data was time intensive and I did some manual classification to train Naïve Bayes. I didn't have time to manually classify all my data so my Naïve Bayes model that I am currently training is not as accurate as it could be. Me manually classifying the data may also introduce bias since I am a fan of Taylor Swift

# System Demo*

To be given on April 10th

# Questions?
# Comments?