

# ***R/NARUTO VS. R/ANIME***

**A PRESENTATION ON CLASSIFICATION MODELS USING  
EARLY 21<sup>ST</sup> CENTURY MACHINE LEARNING TECHNIQUES**

**PRESENTED BY: DANIEL WANG** (OF THE NEO-EARTH RESEARCH COUNCIL)



# 3023 Mars Conference on Pre-World War 3 Humanoid Anthropology

***WELCOME***



# PRESENTATION OUTLINE

## 1. PROBLEM STATEMENT

*Recovered Pre-WWIII data discs of popular 21<sup>st</sup> century website 'reddit'*

## 2. BACKGROUND INFO

*What is anime? Who is Naruto?*

## 3. DISCUSSION OF METHODS

*21<sup>st</sup> Century Machine Learning  
& Classification Methods*

## 4. RESULTS & FINDINGS

*Classification Model Results*

## 5. CONCLUSIONS

*Evaluating Findings*

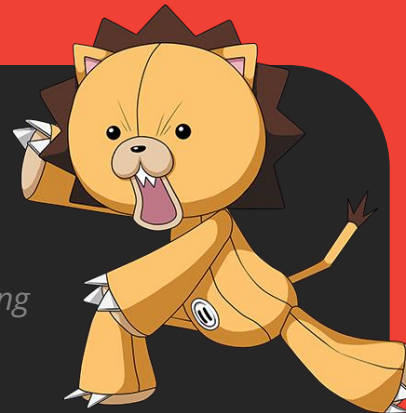




# PROBLEM STATEMENT

- ***Recovered 21<sup>st</sup> century artifacts***

*During a recent expedition to the ruins of alpha-Earth, an exploration team recovered a semi-operable server containing data from the popular 21<sup>st</sup> century website Reddit.*



- ***21<sup>st</sup> century data science techniques***

*Additionally, the team discovered a manual on early 21<sup>st</sup> century data science techniques for classification and machine learning.*

- ***Using 21<sup>st</sup> century methods to analyze 21<sup>st</sup> century data***

*We explored the use of natural language processing and several classification algorithms to differentiate between posts from r/anime vs. those from r/Naruto.*



# ANIME

*Anime is hand-drawn and computer-generated animation originating from Japan. Outside of Japan and in English, anime refers to Japanese animation, and refers specifically to animation produced in Japan. However, in Japan and in Japanese, anime (a term derived from a shortening of the English word animation) describes all animated works, regardless of style or origin. Animation produced outside of Japan with similar style to Japanese animation is commonly referred to as anime-influenced animation.*





*Naruto was an anime/manga series initially airing from 2002-2017\*. The series' popularity coincided with (& helped fuel) an explosion in popularity of anime/manga outside of Japan.*

*It's effects on the genre were significant enough that it can be considered the quintessential shounen\*\* series of the early 21<sup>st</sup> century, upon which many subsequent anime tropes were based.*

*\* Naruto (2002–07)*

*Naruto: Shippuden (2007–17)*

*\*\* shounen = "boys' comics", targeting an audience of adolescent boys*

**NARUTO**  
-ナルト-



# Methods

- 1. Data scraping via API**
- 2. Tokenizing**
- 3. Lemmatizing**
- 4. Stop Words**
- 5. n-word grams**
- 6. Count vectorization**
- 7. Grid Search**
- 8. Confusion Matrices**

# Classification Methods

- 1. Naïve Bayes** (Using Standard CountVectorizer)
- 2. NB w/TfidfVectorizer**  
(Term Frequency-Inverse Document Frequency Vectorizer)
- 3. Random Forest**
- 4. K Nearest Neighbors**



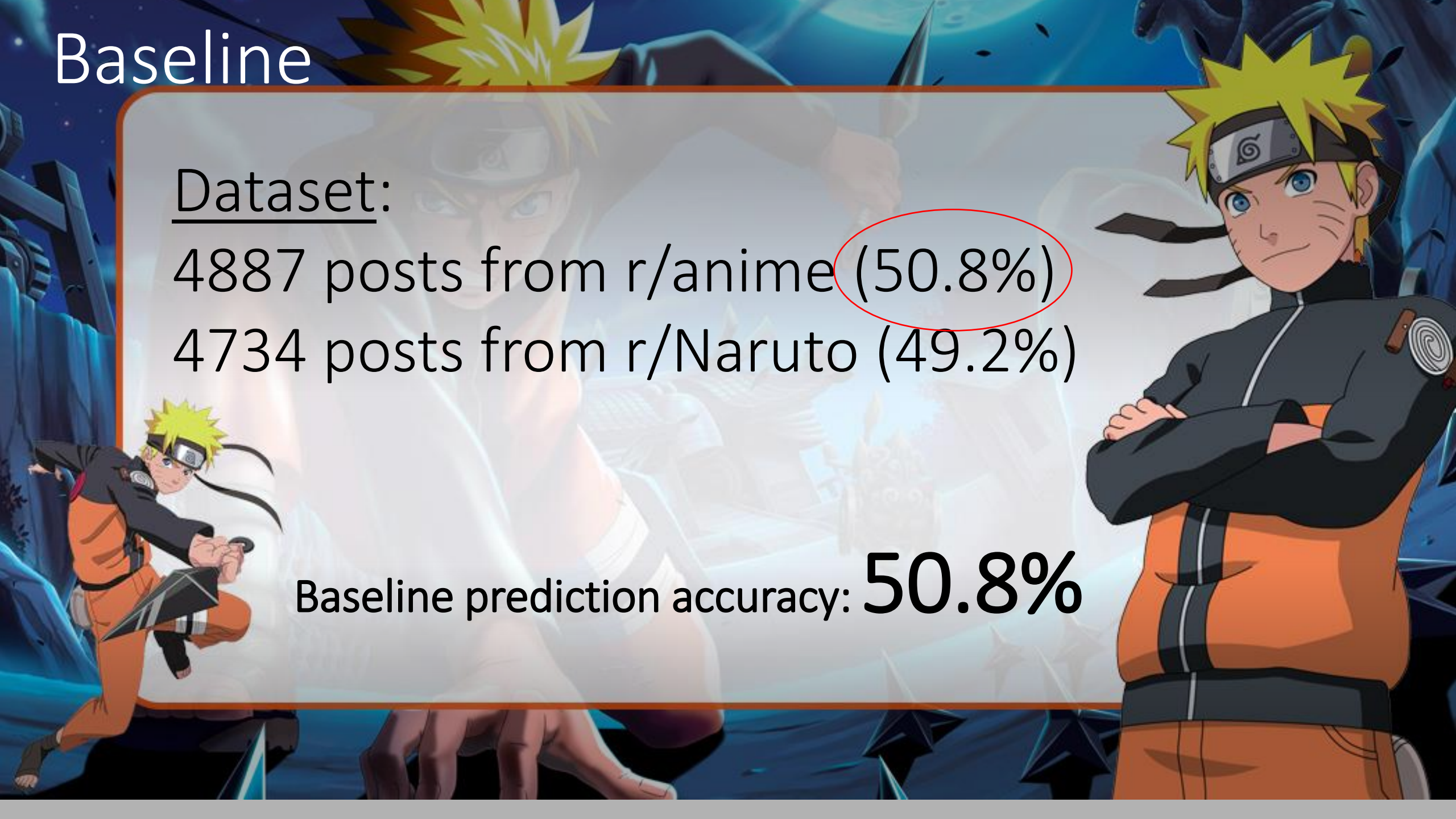
# Baseline

Dataset:

4887 posts from r/anime (50.8%)

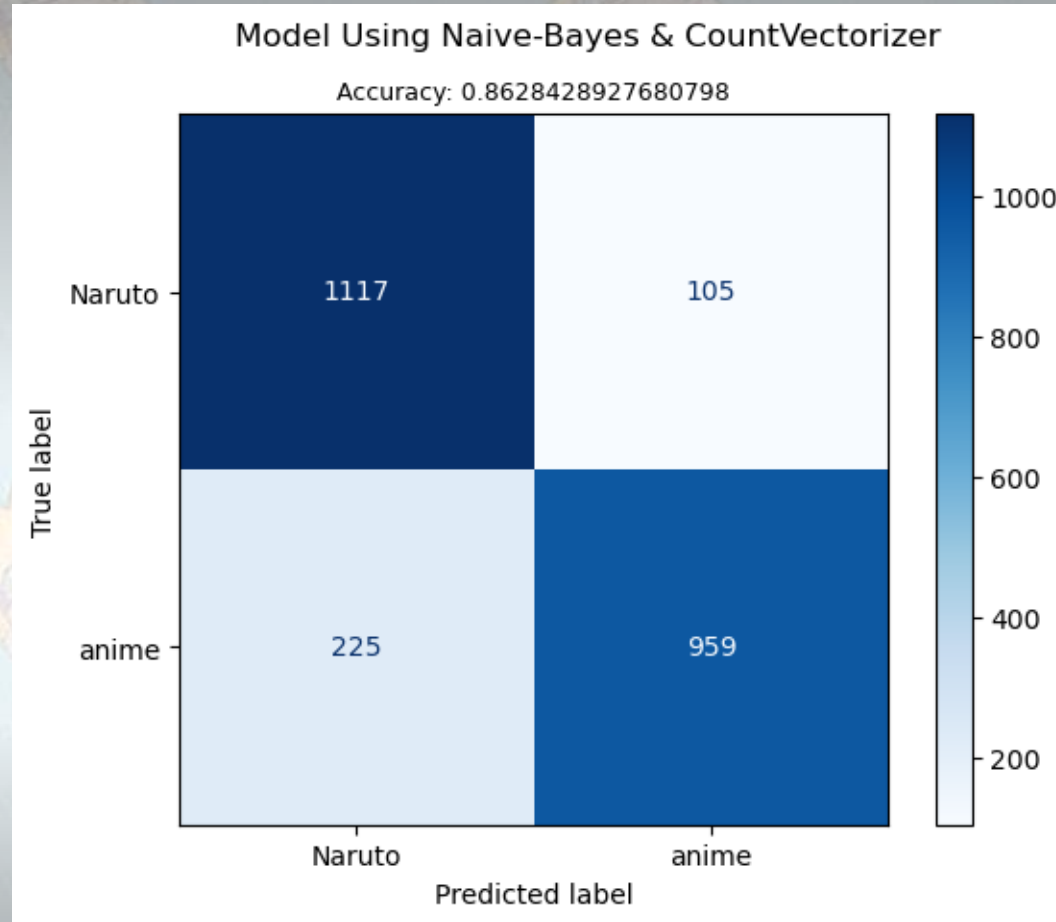
4734 posts from r/Naruto (49.2%)

Baseline prediction accuracy: **50.8%**

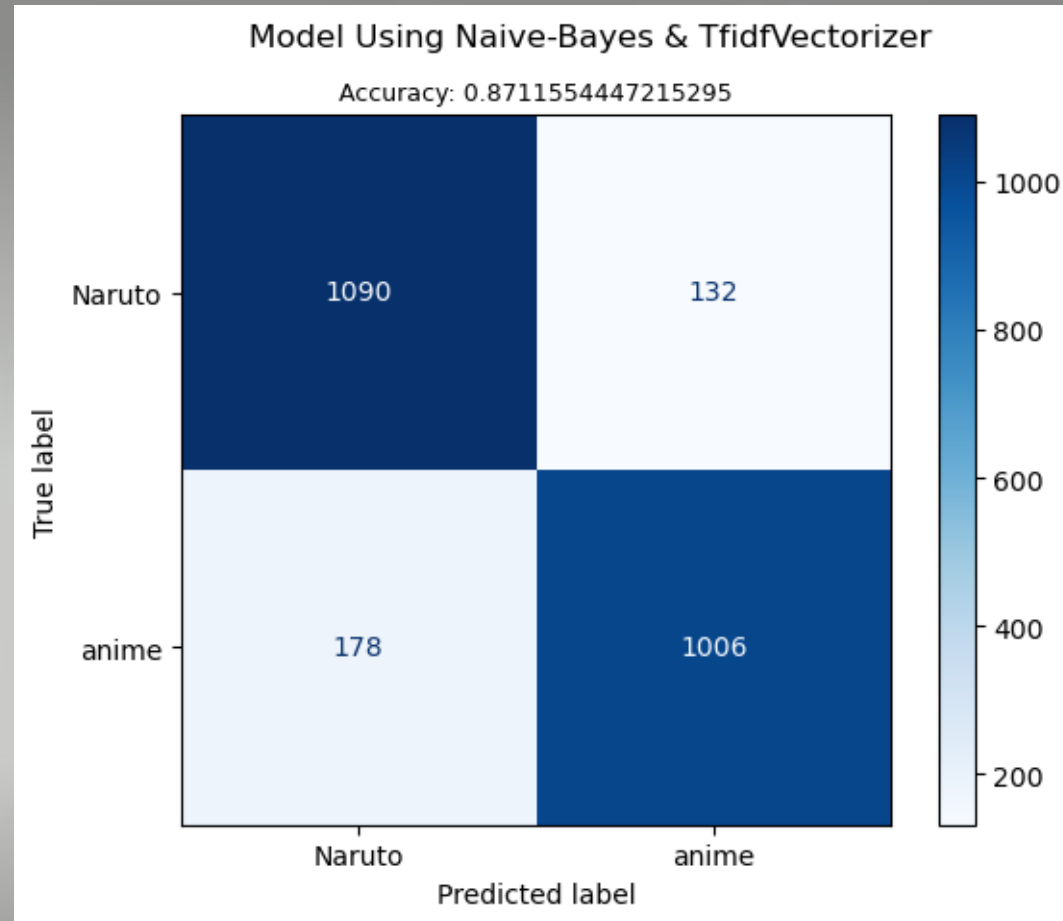




# Naïve Bayes (w/ Standard CountVectorizer)

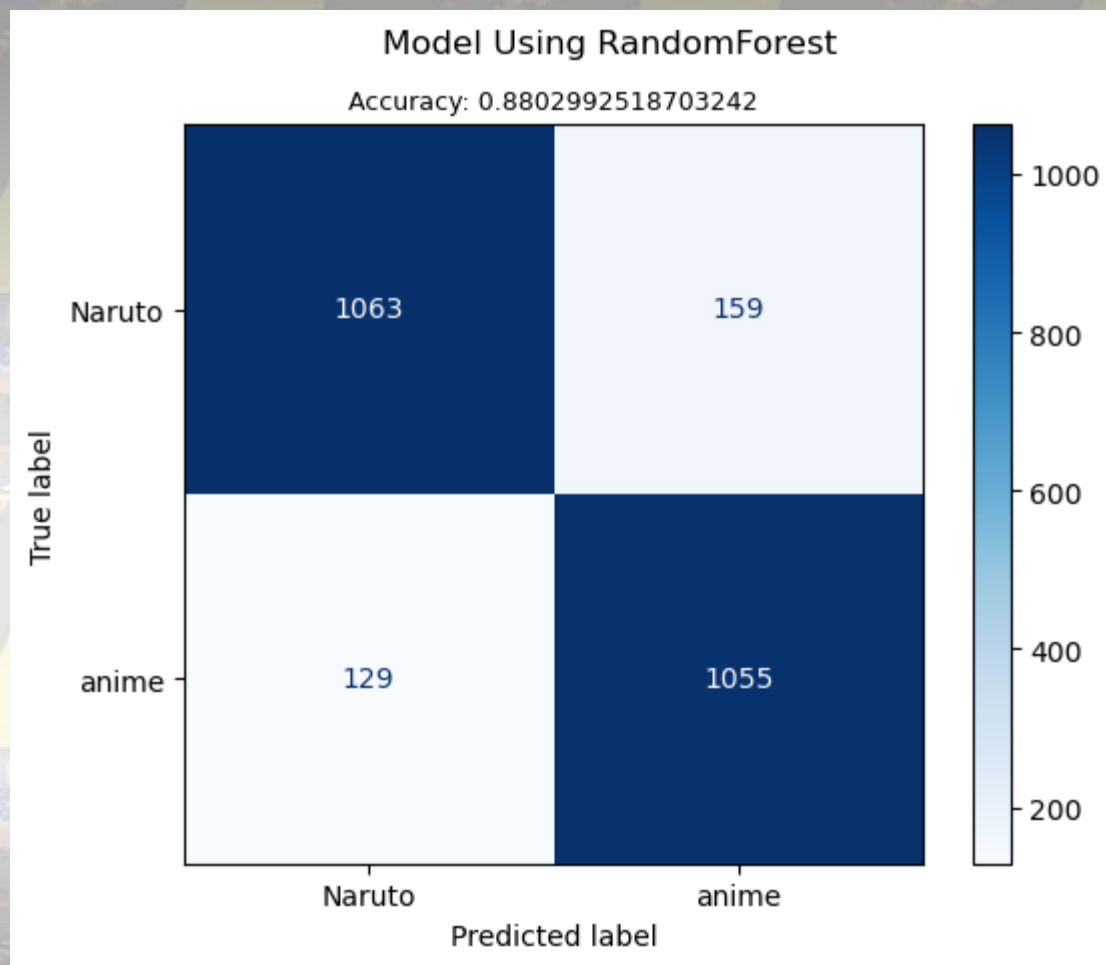


# NB w/ TfidfVectorizer



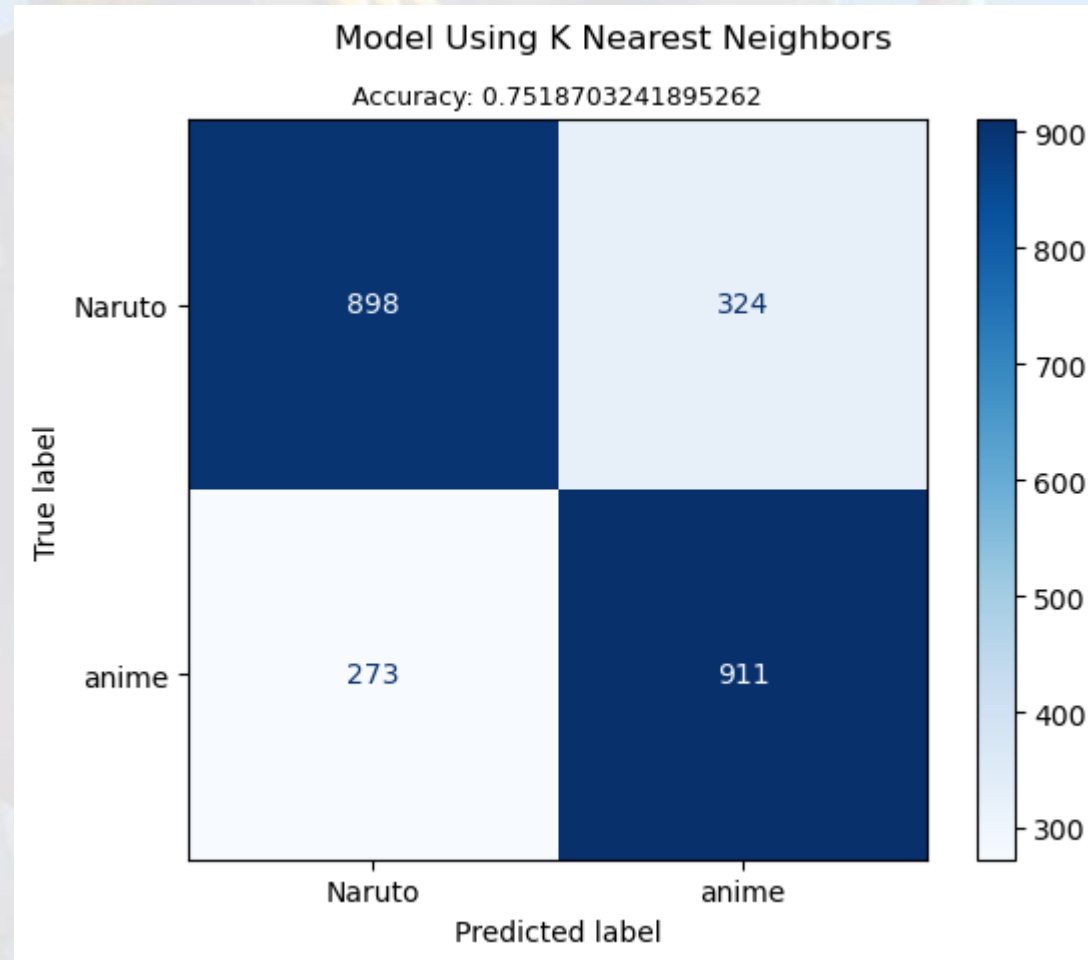


# Random Forest





# K Nearest Neighbors

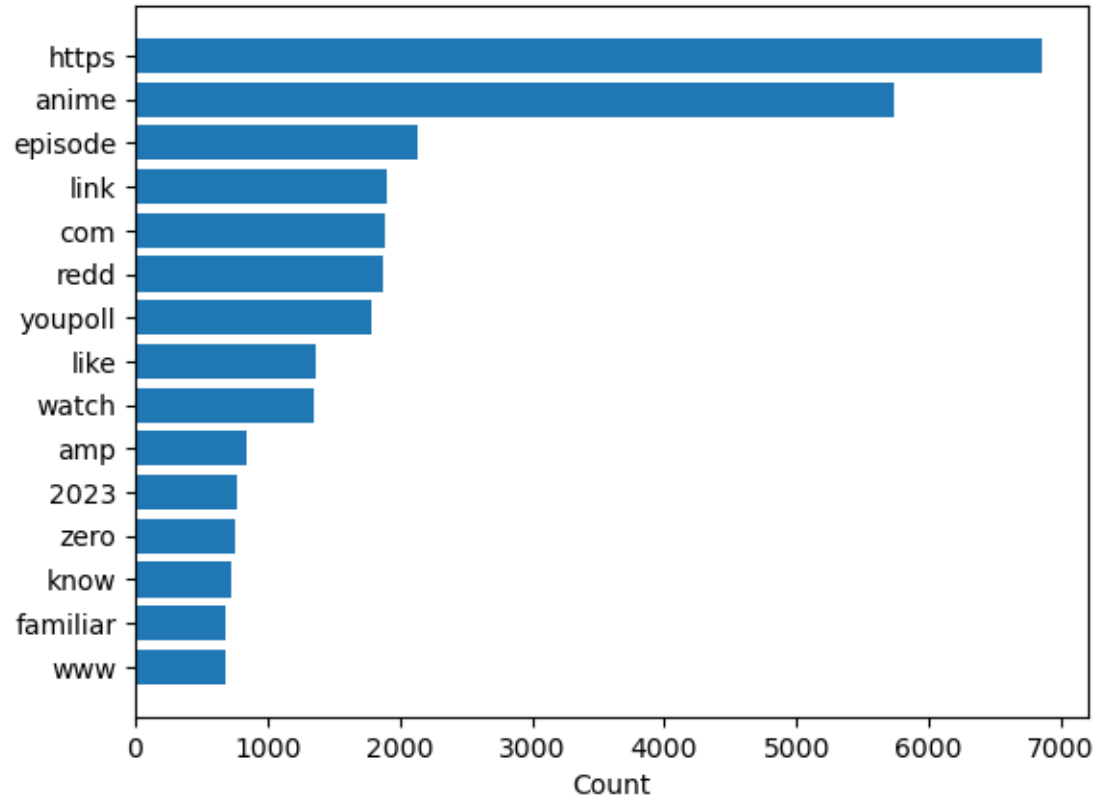


# Final Rankings

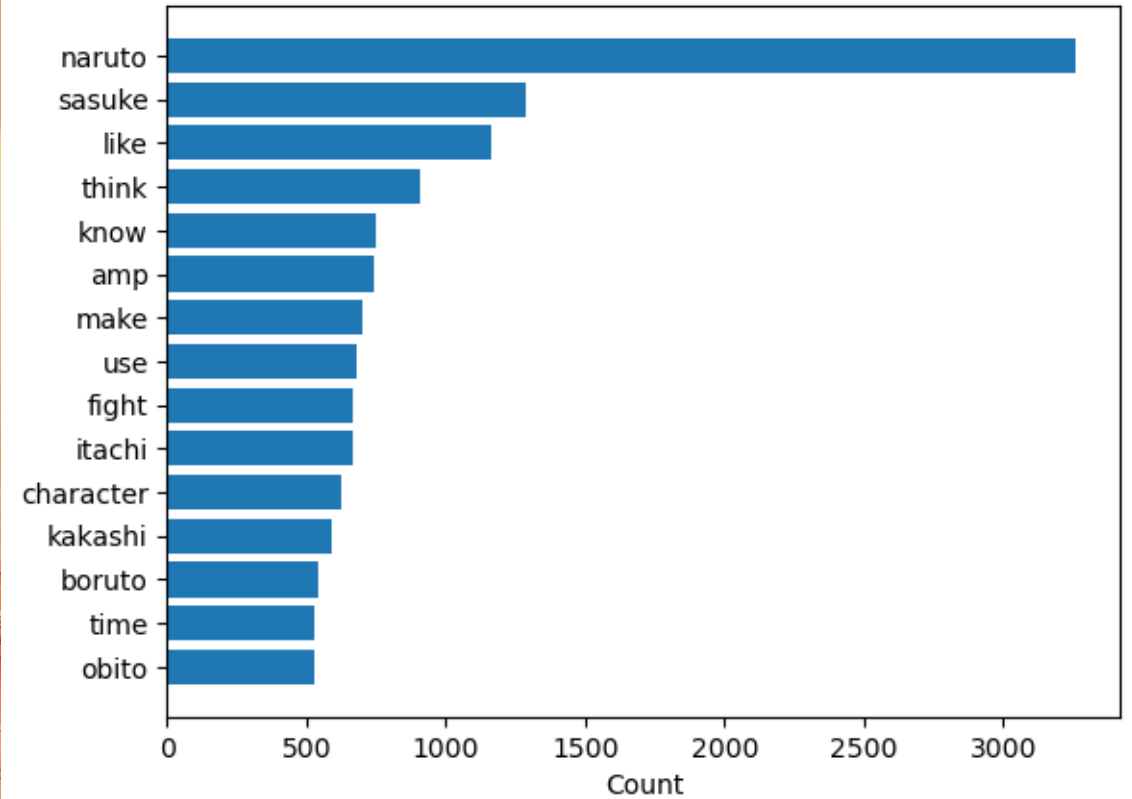
<u>Rank</u>	<u>Algorithm</u>	<u>Accuracy</u>
<b>1.</b>	<b>Random Forest</b>	<b>88.03%</b>
<b>2.</b>	<b>TfidfVectorizer</b>	<b>87.12%</b>
<b>3.</b>	<b>Naive Bayes</b>	<b>86.28%</b>
<b>4.</b>	<b>K Nearest Neighbors</b>	<b>75.19%</b>

# Most common words

15 most common words in anime subset

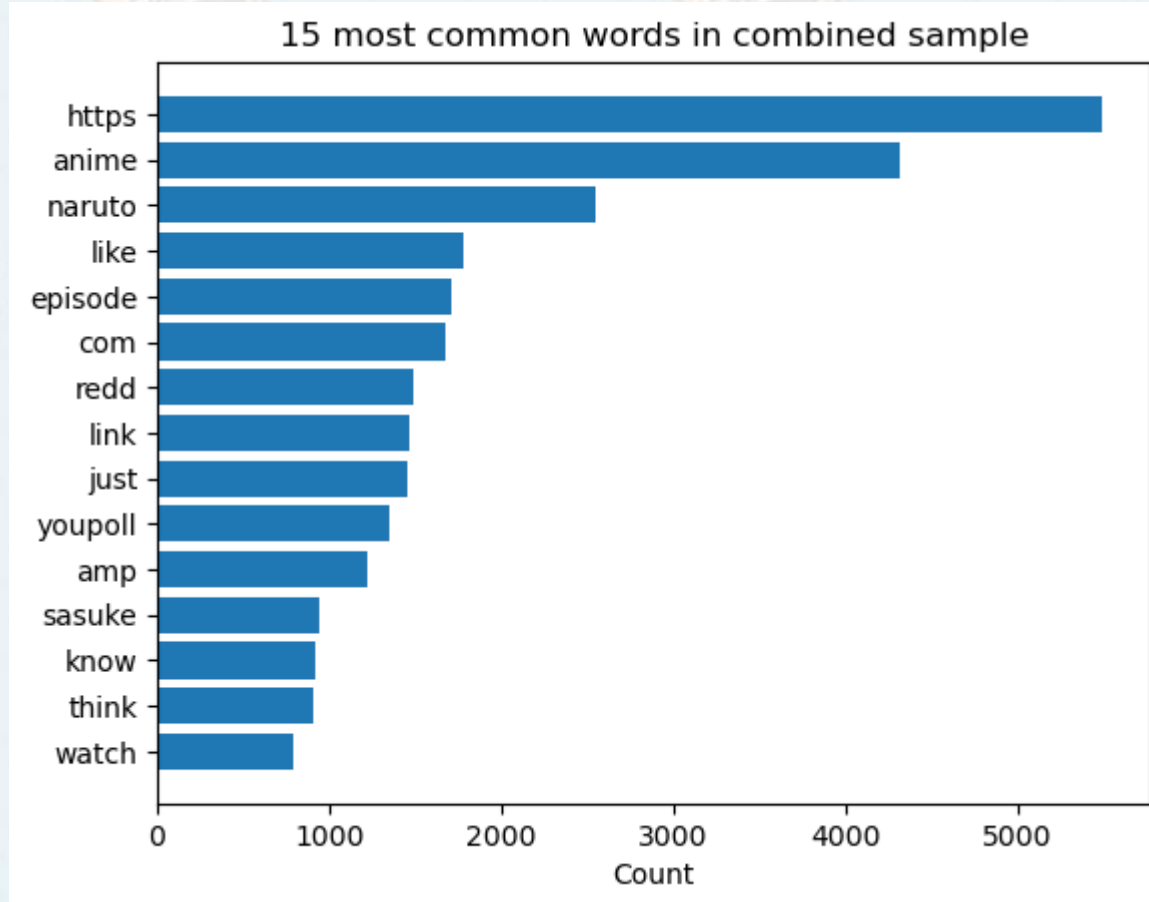


15 most common words in Naruto subset



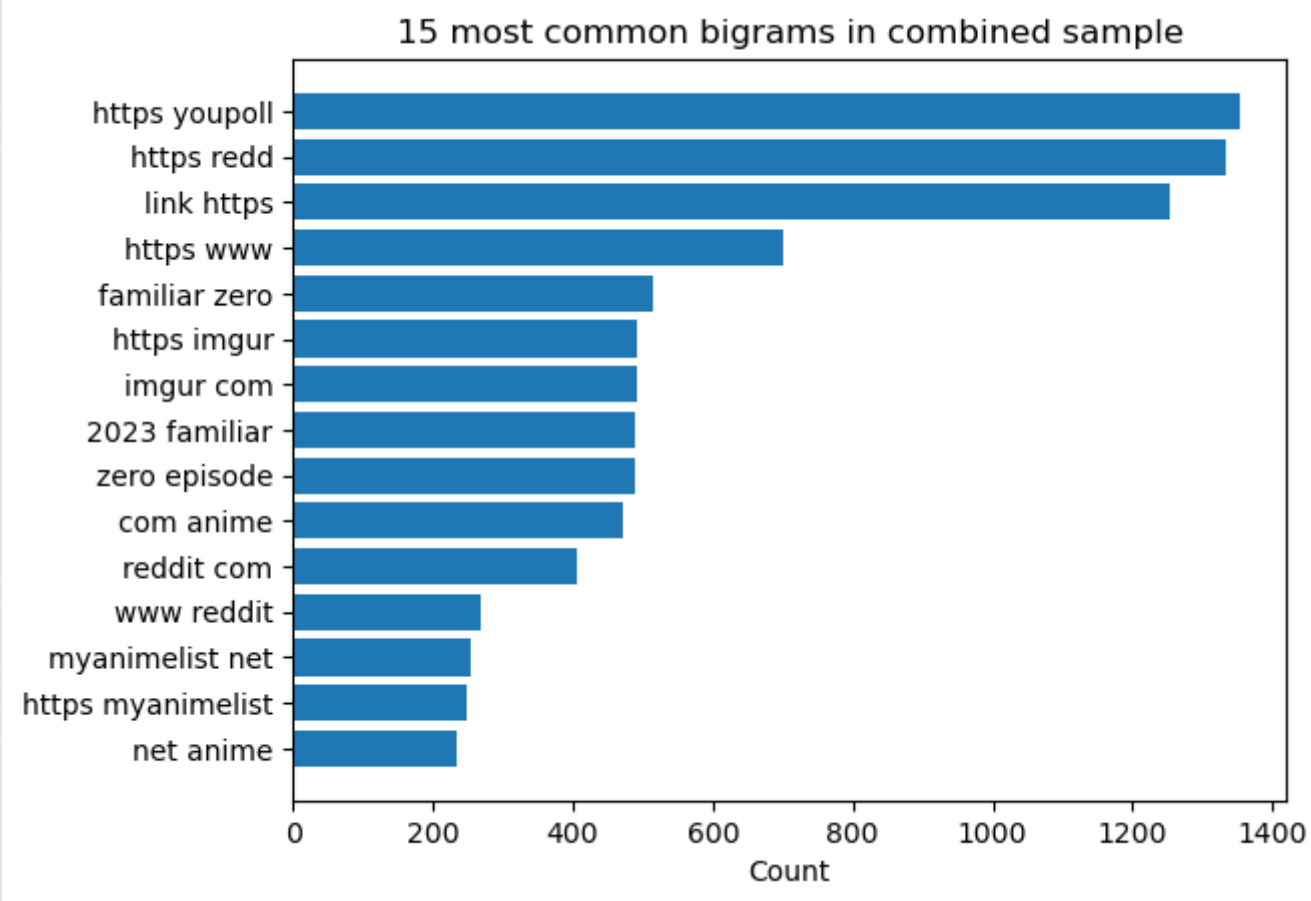


# Most common words

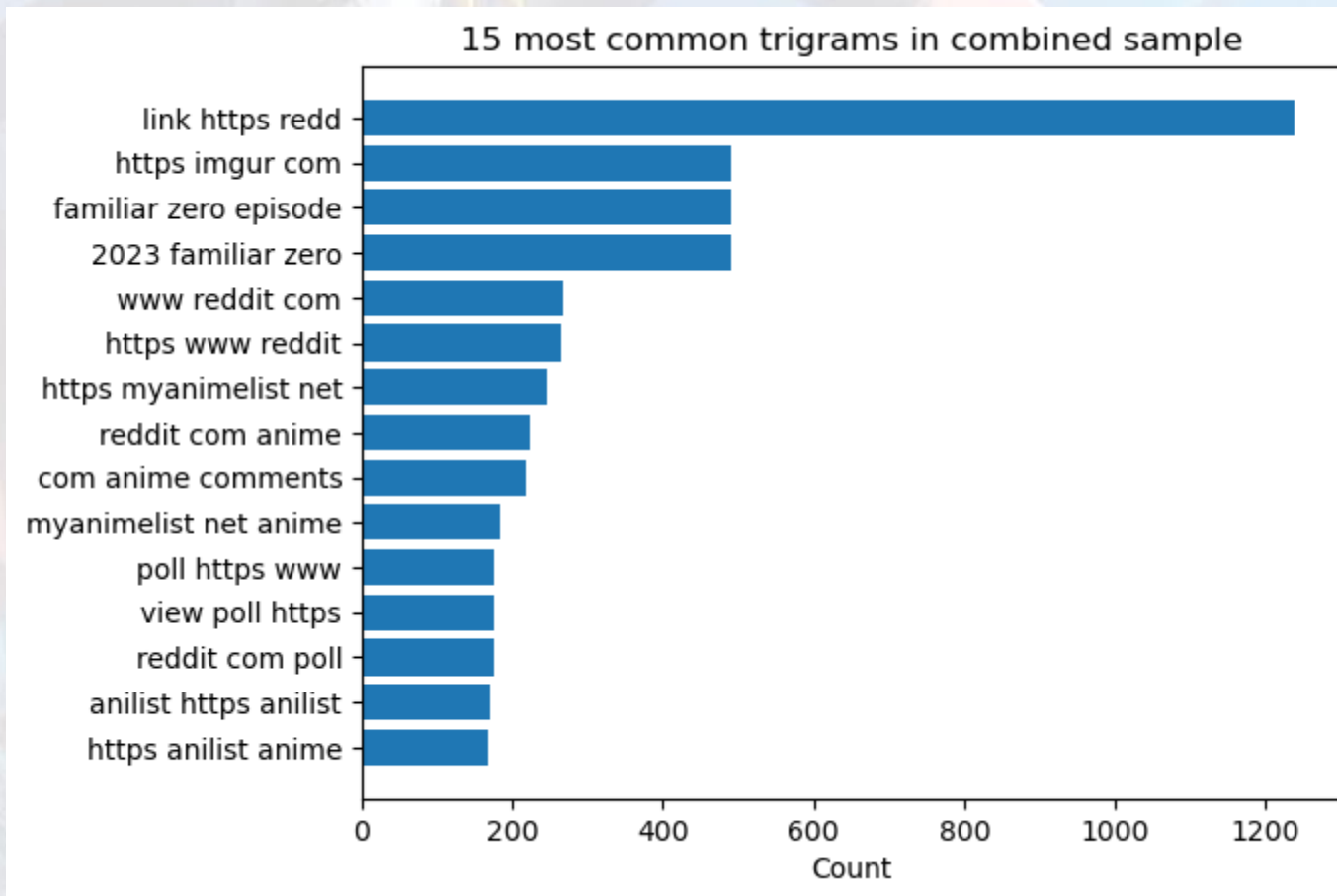




# Most common bigrams



# Most common trigrams





The background of the slide is a collage of various anime panels. In the top left, a character in a dark suit walks. In the top center, a character is seen from behind with a speech bubble saying "...IN A FLASH.". In the top right, a character with a wide, toothy grin has a speech bubble saying "I AM". In the bottom left, a character's face is shown in profile with a speech bubble saying "ARE YOU GOING TO KILL ME RIGHT HERE?". Next to it, another character says "...WHAT?". In the bottom center, a character is falling or jumping. In the bottom right, a character is cooking with a speech bubble saying "FROM TODAY ON", and another character says "WE'RE BROTHERS!!". A speech bubble from a character on the far right says "YOU KNOW, IF WE SHARE A DRINK TOGETHER WE'LL BE BROTHERS!".

# CONCLUSIONS

*The Random Forest (with unlimited depth, single-word n-grams only, and without using stop words) narrowly outperformed its competitors with an accuracy of 88.03%.*

*Based on our findings, all techniques explored yielded better classification results than our baseline. The text discussions from r/anime CAN be reliably differentiated from those of r/Naruto*

*The results of our study provide an interesting window into early 21<sup>st</sup> century alpha-Earth pop culture as well as the contemporary data science techniques of the era.*



A young man with spiky brown hair and a young woman with dark hair in a ponytail are shown from the waist up, looking upwards. They are wearing school uniforms. The background is a bright blue sky filled with fluffy white clouds. The text "THANK YOU!" is written in large, bold, black, italicized capital letters across the center of the image. Below it, the text "QUESTIONS/COMMENTS?" is written in smaller, bold, black, italicized capital letters.

# ***THANK YOU!***

***QUESTIONS/COMMENTS?***



# ***CREDITS***

Data Scraping: [Pushshift Reddit API](#)

Main Slides Template: [Slide Chef](#)

Additional Graphics: Drevio [[1](#)]  
imagemlegal.com.br [[1](#), [2](#), [3](#)]  
fazendoanossafesta.com.br [[1](#)]

