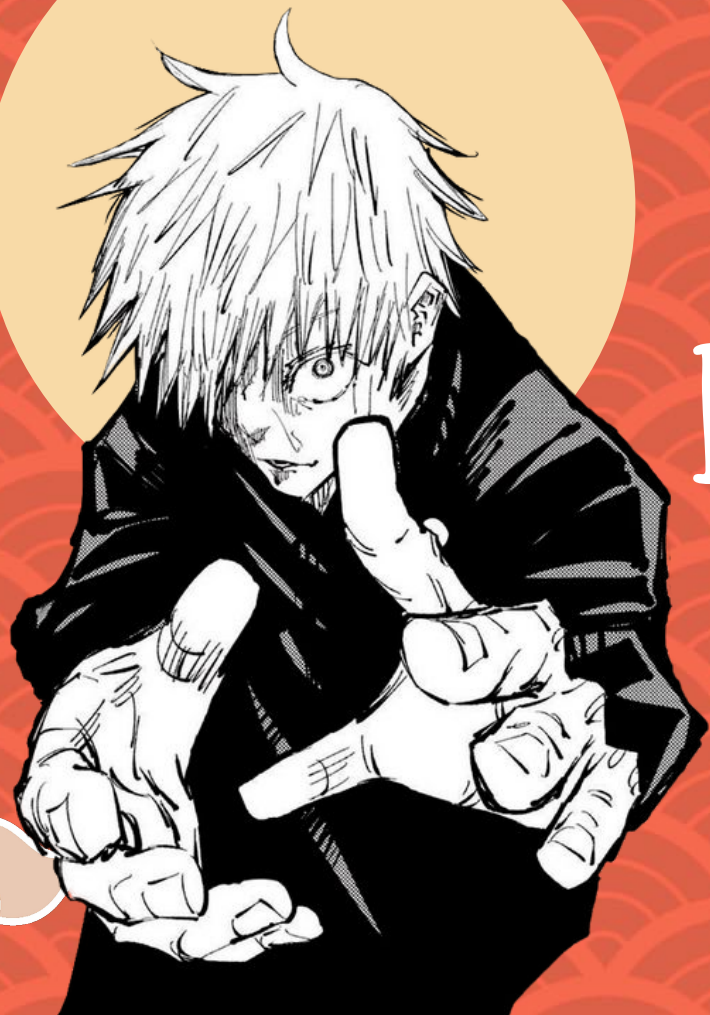


# DATA DRIVEN INSIGHTS ON ANIME

BY CHANDINI & JOEL





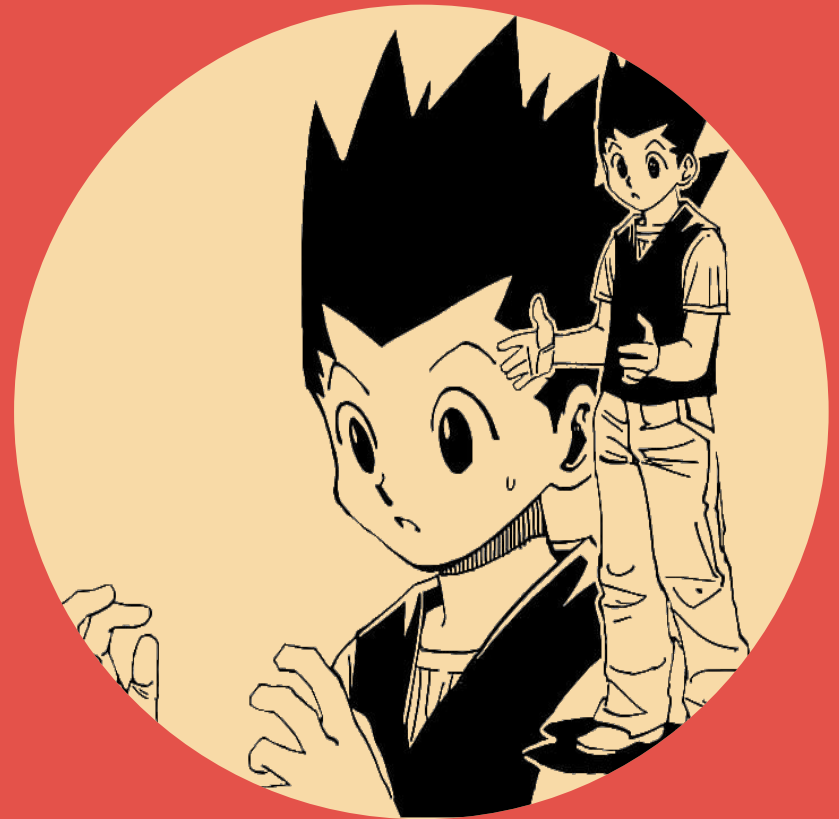
01

# Introduction



# Motivation

At a customs event, Chani and Joel were scared to be alone. So when Joel saw Chani wearing an anime tee, they clung to each other and have been besties ever since. Anime has been a huge part of the friendship.



Gon is a character from Hunter x Hunter, Chandini's favorite anime!



# Our Question

Which factor has the  
greatest impact on the  
rating of an anime?



# The Data

Anime dataset includes rankings of 20k+ anime

Main Features :

- Genre
- Episode length (short, medium, long)
- Licensor
- Source (original, manga, game etc)
- Rating (PG-13, PG, R etc).

We're hoping to determine which feature  
is most informative of an anime being a  
“bop” or a “flop”, meaning it has a score  
above or below a certain threshold



# Data Cleaning

Name	English na	Other na	Score	Genres	Synopsis	Type	Episodes	Aired	Premiere	Status	Producers	Licensors	Studios	Source	Duration	Rating	Rank	Popularity	Favorites	Scored By	Mem
Cowboy B	Cowboy B	ã,«ã, 'ãfœ	8.75	Action, Av	Crime is	TV	26	Apr 3, 199	spring 199	Finished	A Bandai Vi	Funimatic	Sunrise	Original	24 min pe	R - 17+ (vi	41	43	78525	914193	177
Cowboy B	Cowboy B	ã,«ã, 'ãfœ	8.38	Action, Sc	Another	Movie	1	1-Sep-01	UNKNOWN	Finished	A Sunrise, B	Sony Pictu	Bones	Original	1 hr 55 mi	R - 17+ (vi	189	602	1448	206248	36
Trigun	Trigun	ãf~ãf©ã,ã	8.22	Action, Ac	Vash the	TV	26	Apr 1, 199	spring 199	Finished	A Victor Ent	Funimatic	Madhouse	Manga	24 min pe	PG-13 - Te	328	246	15035	356739	72
Witch Hur	Witch Hur	Witch Hur	7.25	Action, Dr	Robin	TV	26	Jul 3, 200	summer 2	Finished	A Bandai Vi	Funimatic	Sunrise	Original	25 min pe	PG-13 - Te	2764	1795	613	42829	11

Episodes	Licensors	Source	Rating	Action	Scifi	Adventure	Drama	Mystery	Supernatu	Fantasy	Sports	Comedy	Romance	Suspense	Ecchi	Horror	Slice of Life	Girls Love	Boys Love	class
Short	Funimatio	Original	R-17+	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
Short	Sony_Pictu	Original	R-17+	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
Short	Funimatio	Manga	PG-13_Tee	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1
Short	Funimatio	Original	PG-13_Tee	1	0	0	1	1	1	0	0	0	0	0	0	0	0	0	0	0

- Dataset originally had 24 different feature columns
- Focused on 5 main features: genre, episode length, licensor, source and rating
- Removed all the rows of data that included unknown values
- Since each anime is classified under multiple genres, separated the genre column into individual columns with 1 or 0 as the value if it include it
- Converted the score column as 'class', since it's what we're trying to predict and set the values as either 1 or 0 (bop or flop)





02

# Methods



# Best Feature: Entropy/ Accuracy

- Found the best feature by info gain and classification accuracy for 3 thresholds for being a bop; 8/10, 7.5/10, and 6/10.

## Naive Bayes

- Naive Bayes was also run for each threshold



Naruto is a classic anime many beginners start with!





03

**Results**



# Information Gain

**Bop or Flop Parameter set to 6**

**Best Info Gain: 0.04257198126204281**

**Best Feature: Licensors**

**Bop or Flop Parameter set to 7.5**

**Best Info Gain: 0.06017373441896101**

**Best Feature: Licensors**

**Bop or Flop Parameter set to 8**

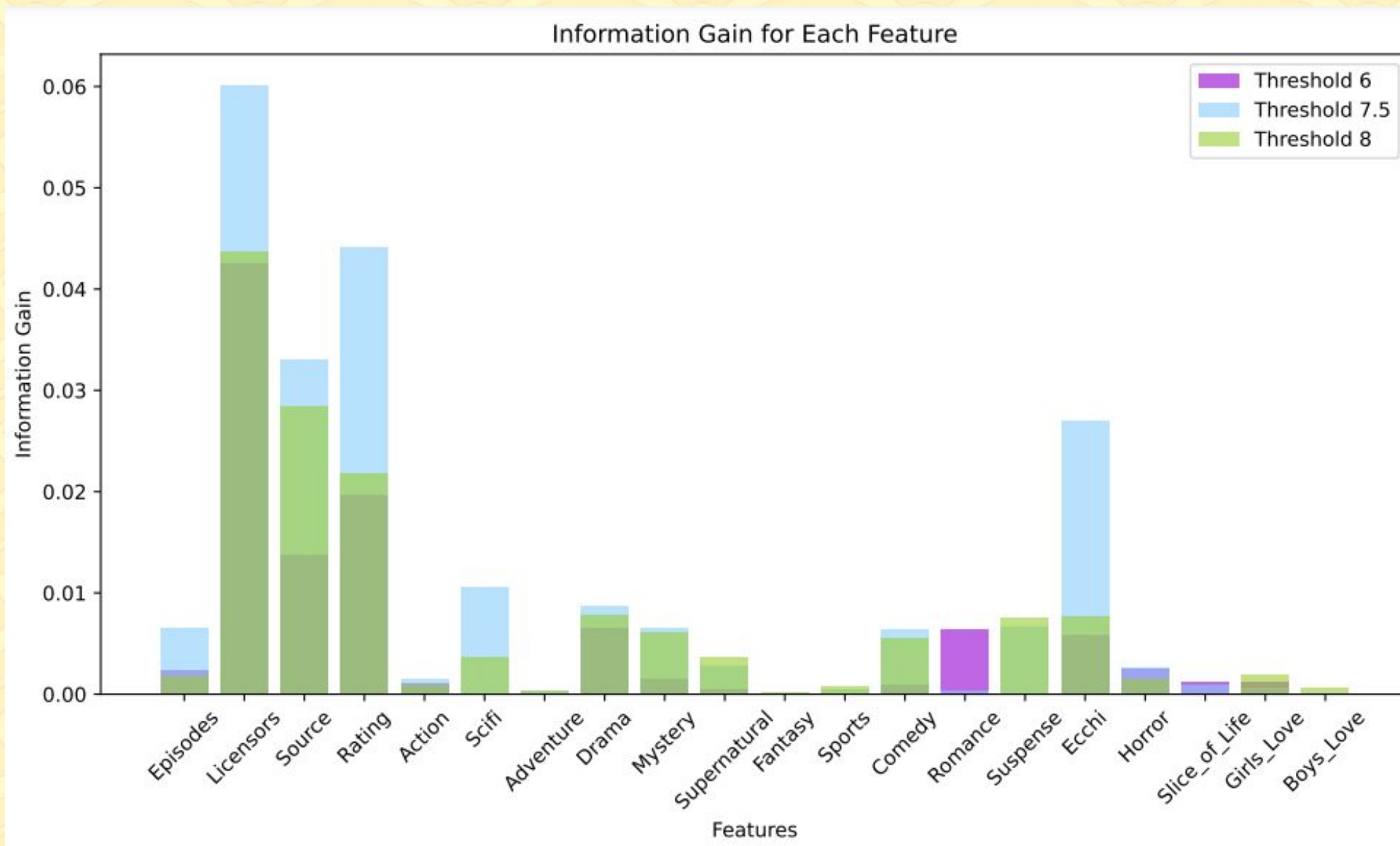
**Best Info Gain: 0.04365848251052562**

**Best Feature: Licensors**





# Info Gain Comparison





# Classification Accuracy

6/10: Best feature was Episode length

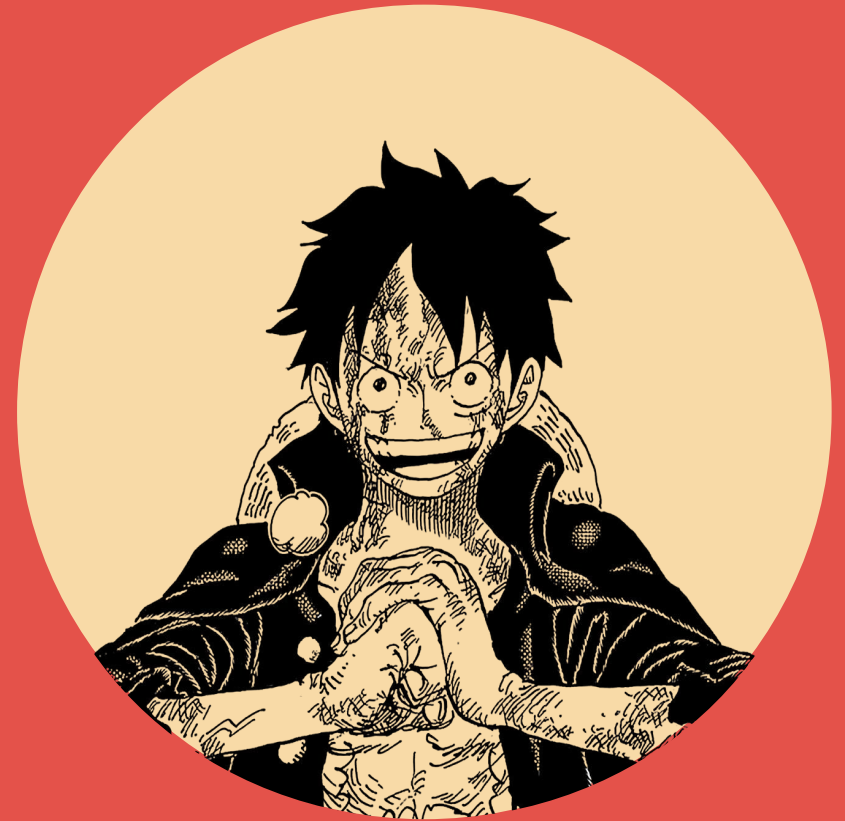
- Info Gain BF was Licensers

7.5/10: Best feature was Licensors

- Info Gain BF was the same

8/10: Best feature was Licensor

- Info Gin BF was the same



Luffy is from the top anime in the world, One Piece! It's also Joel's favorite



# Naive Bayes Results

6/10

93% Accuracy

0: 198 1: 2364

	Feature	Avg_Log_Prob
	Episodes_Short	-1.947901
	Rating_PG-13_Teens_13_or_older	-2.485326
	Action	-2.697384
	Comedy	-2.888725
	Source_Manga	-2.925987
	...	...
	Licensors_Flatiron_Film_Company	-8.101945
	Licensors_Frontier_Works	-8.101945
	Licensors_bilibili	-8.101945
	Licensors_Mill_Creek_Entertainment	-8.448518
	Source_Web_novel	-8.448518

- True Positives (TP): 591
- True Negatives (TN): 4
- False Positives (FP): 39
- False Negatives (FN): 7

7.5/10

72% Accuracy

0: 1814 1: 748

	Feature	Avg_Log_Prob
	Episodes_Short	-1.953012
	Rating_PG-13_Teens_13_or_older	-2.385665
	Action	-2.758425
	Source_Manga	-2.772713
	Comedy	-2.880632
	...	...
	Licensors_Flatiron_Film_Company	-8.618659
	Licensors_bilibili	-8.618659
	Licensors_Voyager_Entertainment	-8.618659
	Source_Web_novel	-8.965233
	Licensors_Mill_Creek_Entertainment	-8.965233

- True Positives (TP): 61
- True Negatives (TN): 399
- False Positives (FP): 51
- False Negatives (FN): 130

8/10

89% Accuracy

0: 2269 1: 293

	Feature	Avg_Log_Prob
	Episodes_Short	-1.970090
	Rating_PG-13_Teens_13_or_older	-2.436054
	Source_Manga	-2.713243
	Action	-2.785111
	Comedy	-2.991189
	...	...
	Licensors_Geneon_Universal_Entertainment	-8.277257
	Licensors_Ketchup_Entertainment	-8.277257
	Licensors_Muse_Communication	-8.277257
	Source_Web_novel	-8.623831
	Licensors_Mill_Creek_Entertainment	-8.623831

- True Positives (TP): 8
- True Negatives (TN): 561
- False Positives (FP): 13
- False Negatives (FN): 59

- The confusion matrix reveals challenges, particularly in handling false positives and false negatives.
- Features with higher average log probabilities, such as "Episodes\_Short" and "Rating\_PG-13\_Teens\_13\_or\_older" are more indicative of the positive class (bop)



04

# Future Research/ Conclusions





# Limitations

- Class imbalance as seen in Naive Bayes
  - Bias towards majority class
  - Misleading accuracy since it appears high but there's a majority class
- Score
  - subjective threshold for bop/flop
  - nothing rated below 6
  - determined by user voting



Jujutsu Kaisen is a new gen anime that now has the top rated episode of all time!

# Areas for Further Research

- Feature interaction analysis
  - ie: ecchi genre informing rating in info gain
- Use data from viewer count/manga sales rather than viewer rating
- Zoom in on certain features more like specific licensors or genres



Dragonball is the first anime a lot of Americans watched since it was often televised! Its Chani's and Joel's first as well!





Thank you!

