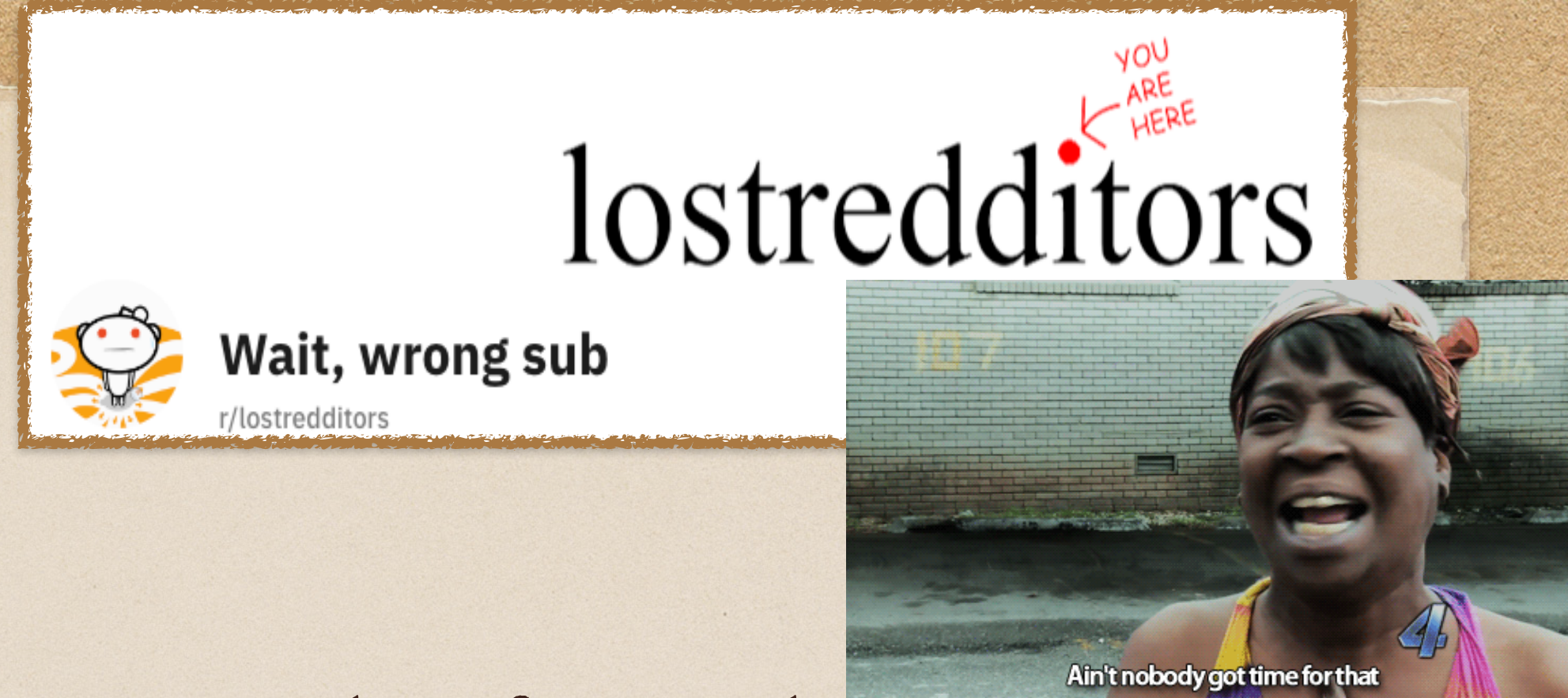# Automate boring stuff

– for r/discworld moderators

Project Aim

Problem:

- Formerly known as "the front page of the internet", Reddit is an online forum platform with over 130,000 communities and over 50 million daily active users worldwide.
- Community Moderators ("Mods") are volunteers who are tasked to maintain the standard of content within each sub, settle disputes, levy rules.
- Lost Redditors is such a prevalent issue there is even a sub for it.

Proposed Solution:

- Utilize Natural Language Processing and Machine Learning models, to predict if a post should belong to r/discworld or r/Cosmere
- This will save time for Mods to concentrate on higher-valued tasks

Why would a lost Redditor get confused?

What is the Discworld:

The Discworld is a flat planet resting on the shoulders of 4 elephants who in turn stand on top of the turtle as it travels through space.
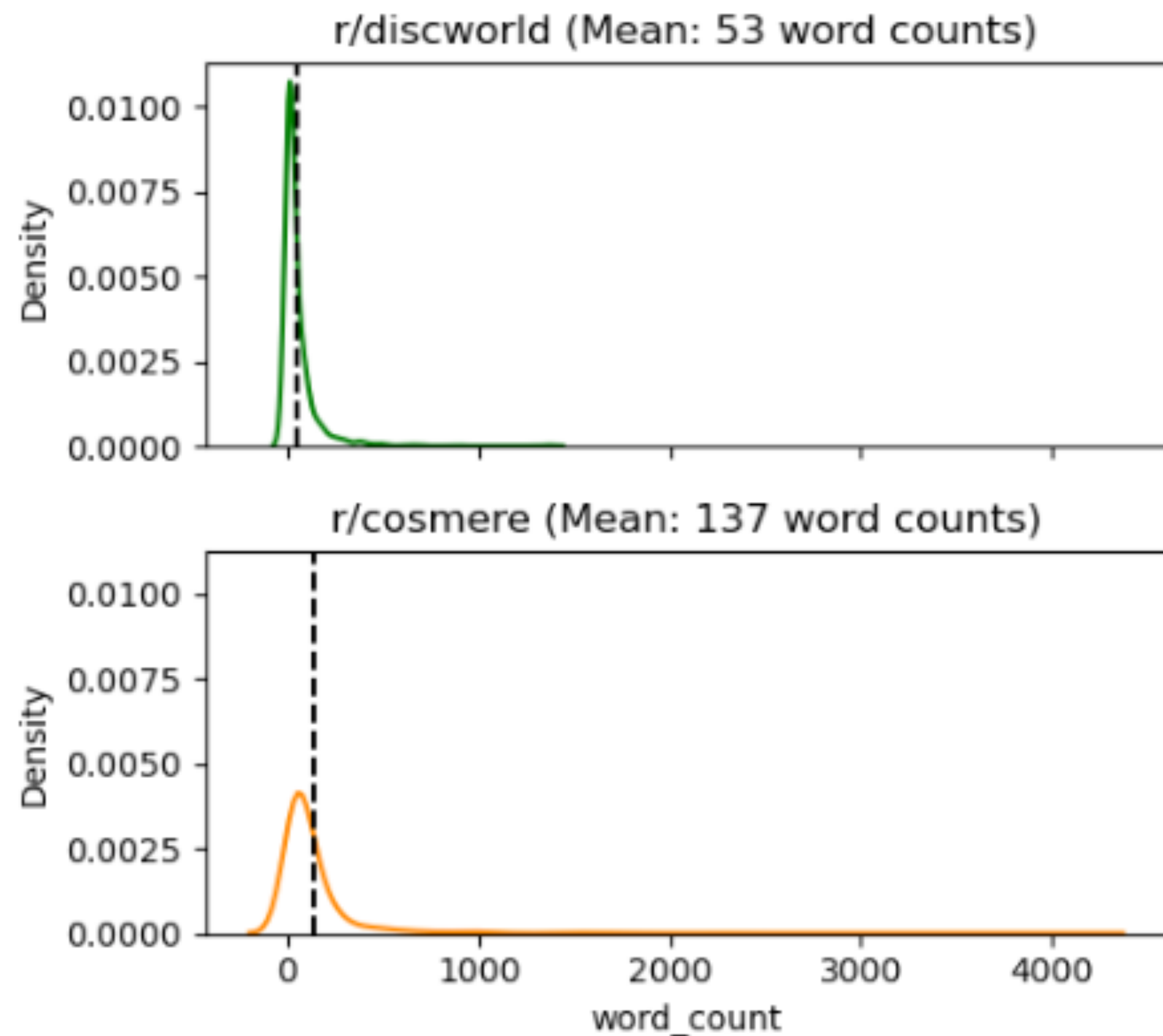
What is the Cosmere Universe:

- Cosmere universe is set after The Shattering, dividing into 16 Shards or Gods. There are 10 known planetary systems. There also exist 3 realms.

- Both series are one of the top high fantasy series, with extensive world building, following multiple characters.
- Large collection: 41 Discworld novels (1983 – 2015), 24 Cosmere novels (2005 -)
- Most books in both series are award winning and top the best-seller lists
- r/discworld has 80K members and r/Cosmere has 110K members
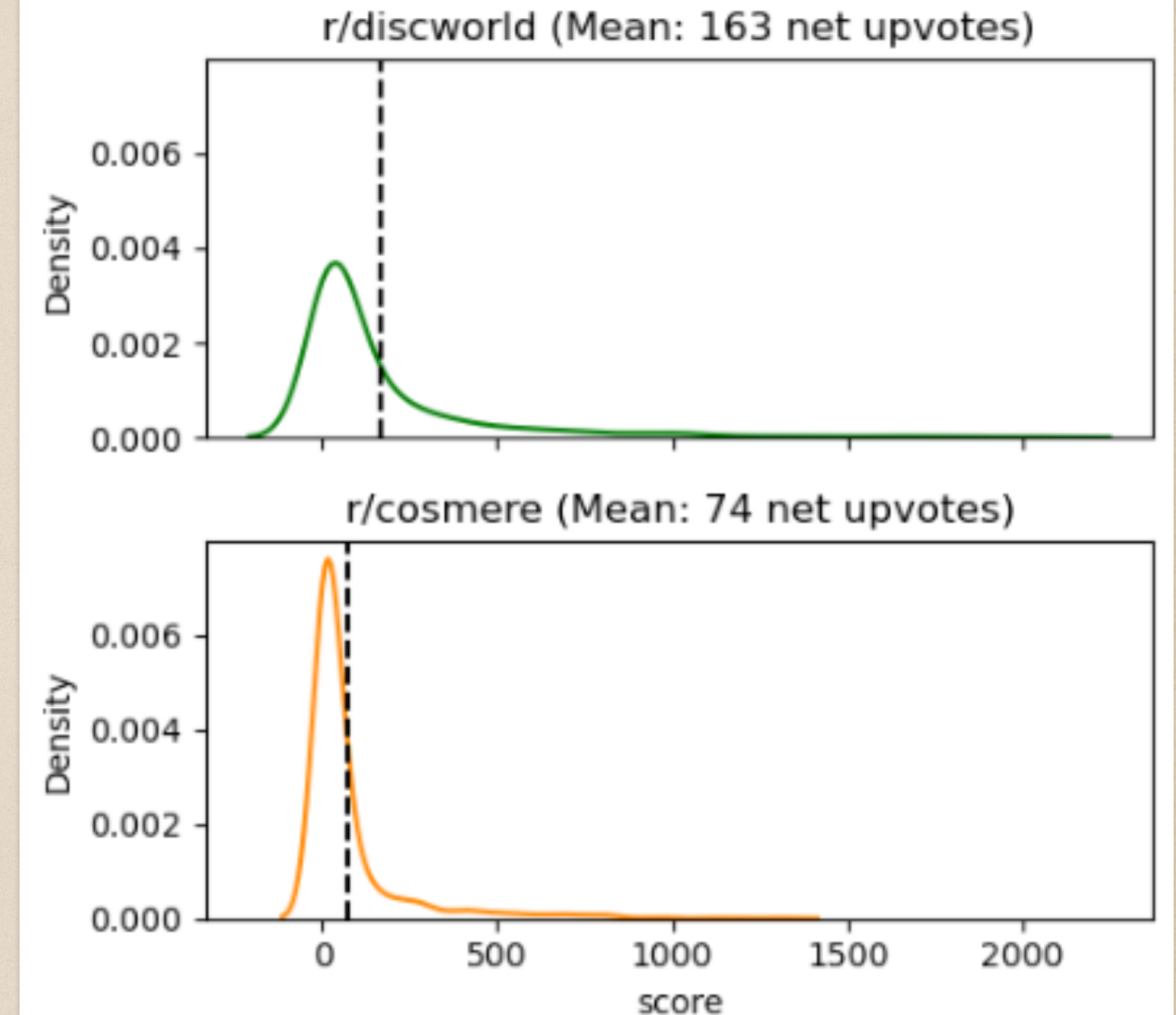
Exploratory Data Analysis
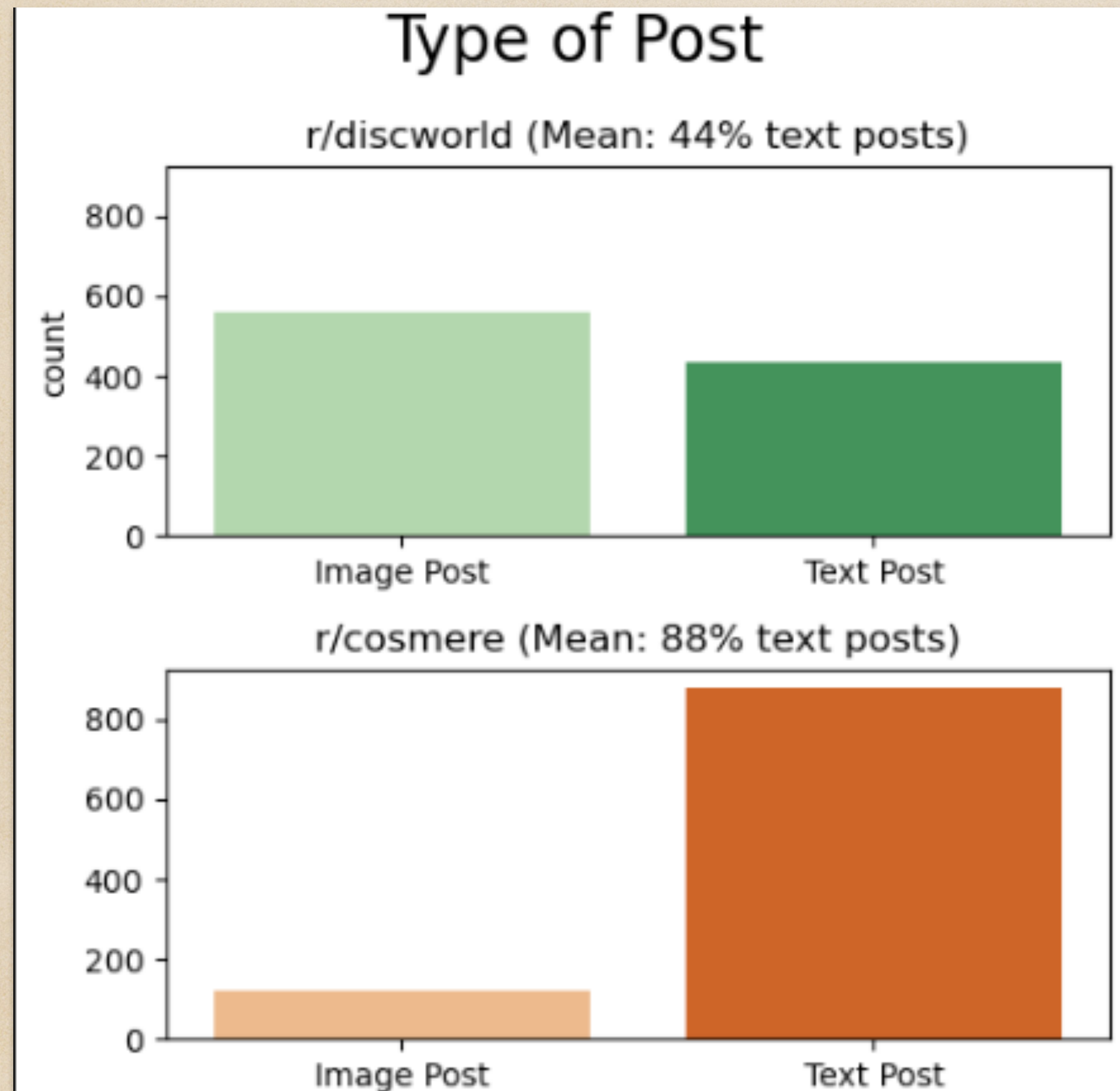
Differences in post features:

Word Counts

r/discworld (Mean: 53 word counts)

r/cosmere (Mean: 137 word counts)

Net Upvotes

r/discworld (Mean: 163 net upvotes)

r/cosmere (Mean: 74 net upvotes)

Subs	Average	Max
r/discworld	53	1,369
r/Cosmere	137	4,172

Subs	Average	Max
r/discworld	163	2,040
r/Cosmere	74	1,300

Differences in post features:

Type of Post

r/discworld (Mean: 44% text posts)

r/cosmere (Mean: 88% text posts)

Type of Post

r/discworld (Mean: 5% spoiler posts)

r/cosmere (Mean: 78% spoiler posts)

Subs	Tex Posts	Image Post
r/discworld	44%	56%
r/Cosmere	88%	12%
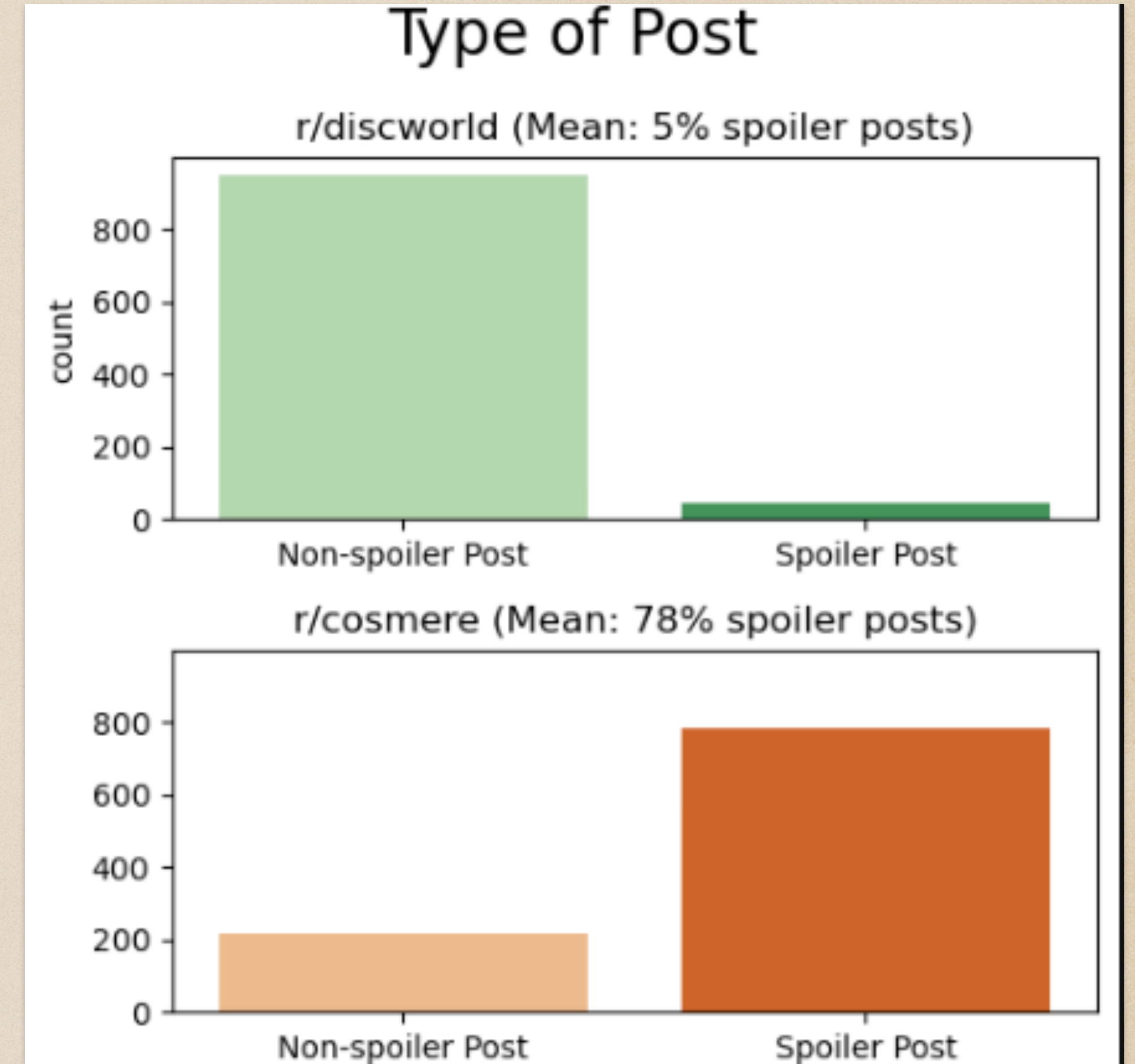
Subs	Non Spoiler	Spoiler
r/discworld	95%	5%
r/Cosmere	22%	78%

Top Words used in posts:

r/discworld   r/Cosmere

Top Words used in posts:

good
discworld
death
get
pratchett
time
man
year
find
today
terry
go

ve
read
not
know
like
do
think
book

mistborn
shard
era
question
stormlight
tress
way
spoiler
metal
want
hoid
cosmere

r/discworld
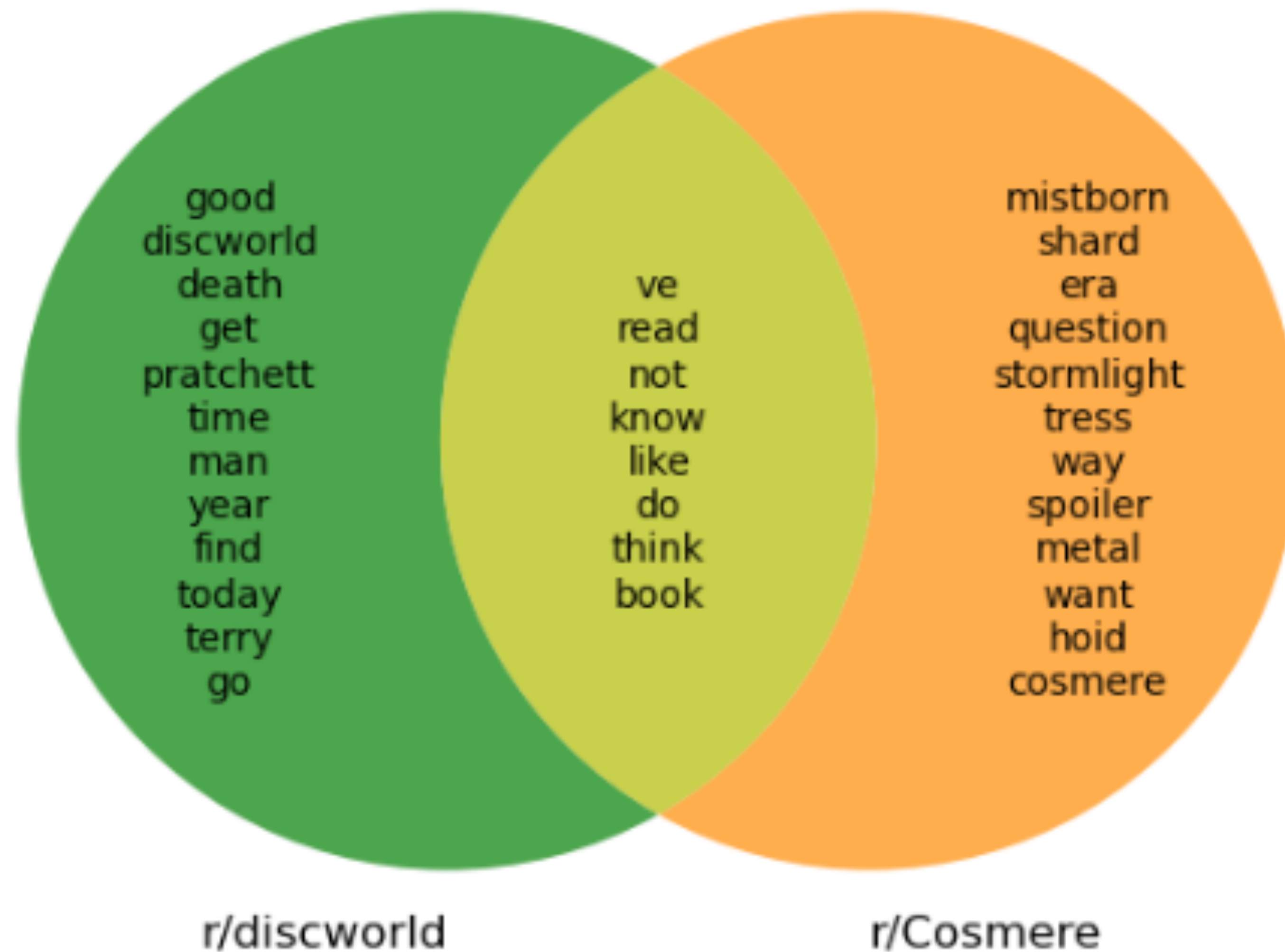
r/Cosmere

Top Bi-grams used in posts:
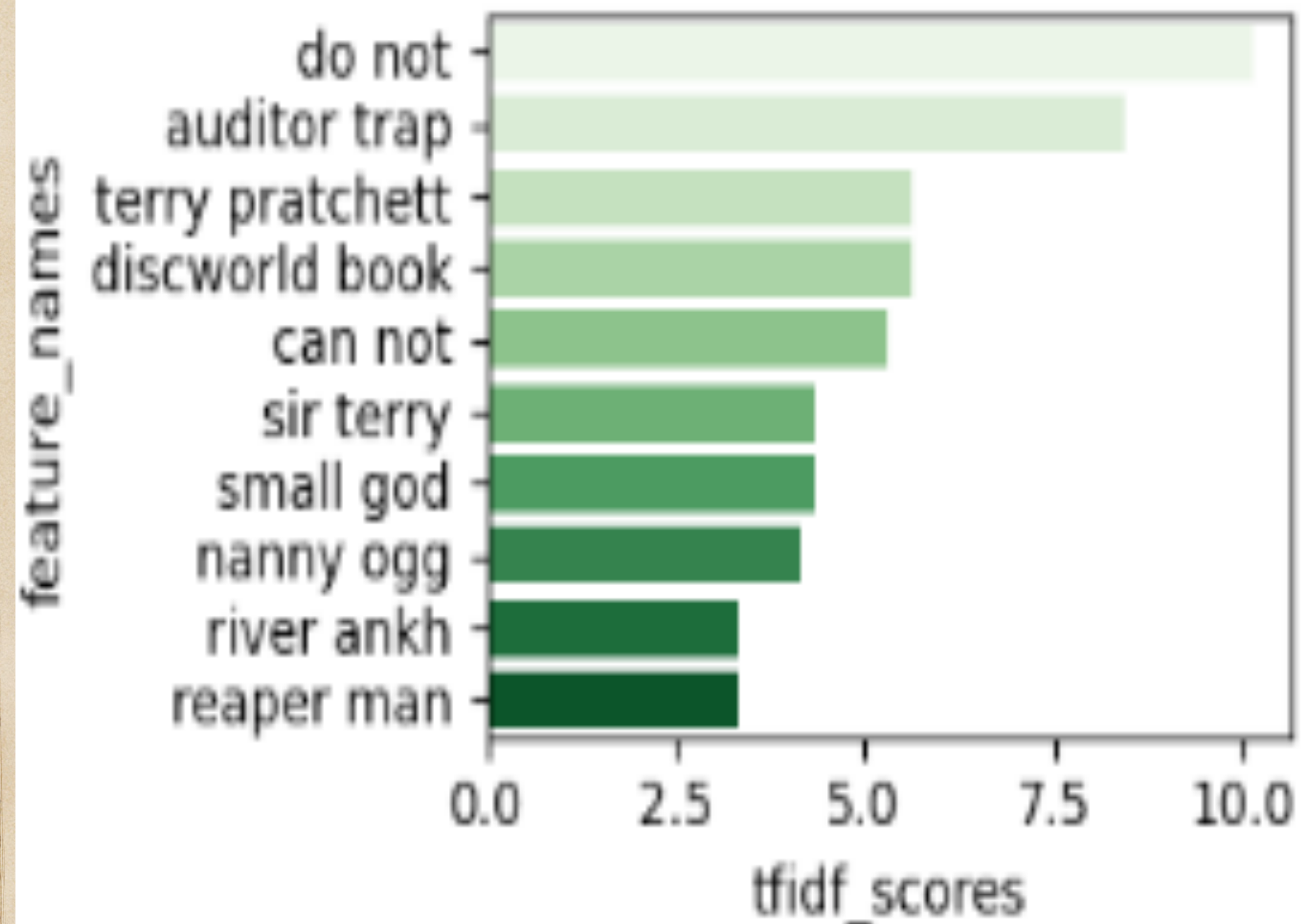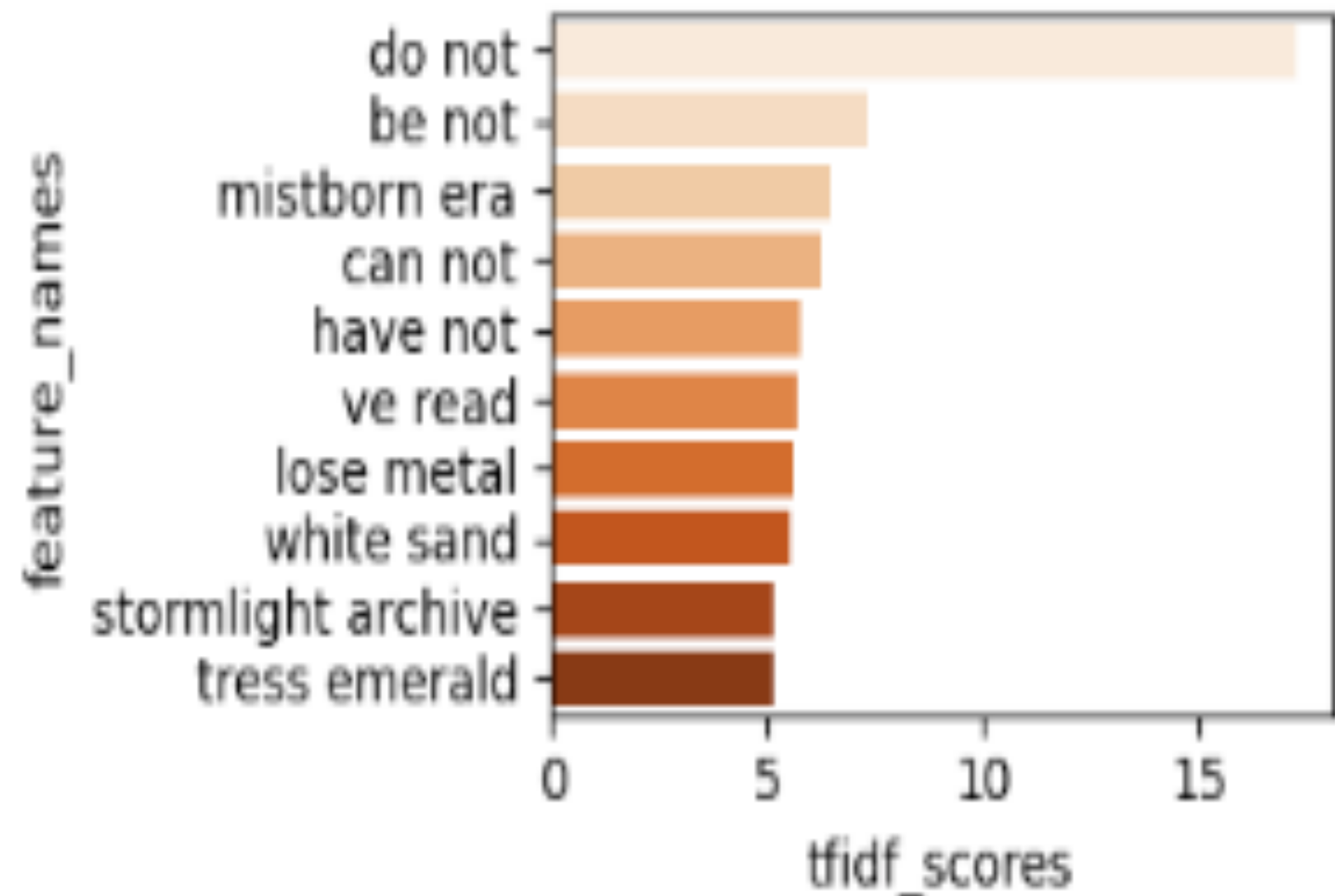
Sentiments of posts:

Sentiment Score

r/discworld (Mean: 0.22 sentiment score)

r/cosmere (Mean: 0.27 sentiment score)
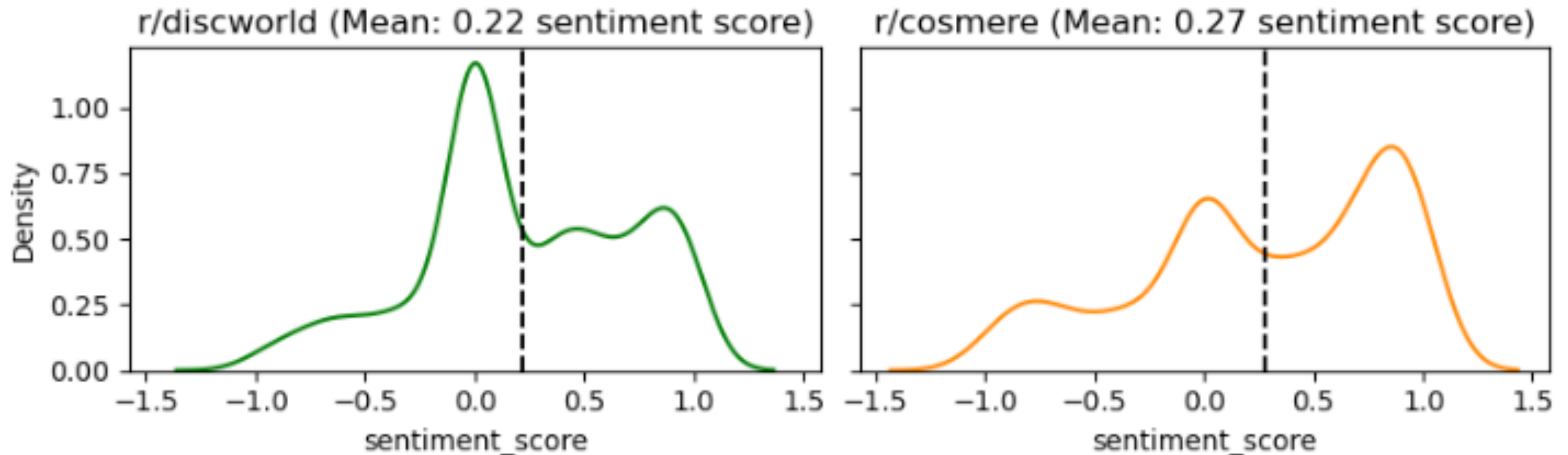
r/discworld	Proportion
Positive Posts	50%
Neutral Posts	31%
Negative Posts	18%

r/Cosmere	Proportion
Positive Posts	60%
Neutral Posts	23%
Negative Posts	17%

Modelling

++?????++ Out of Cheese Error.

Numerical Features:

- Word Count
- Sentiment Score

Categorical Features:

- is_text post tag
- Is_spoiler post tag

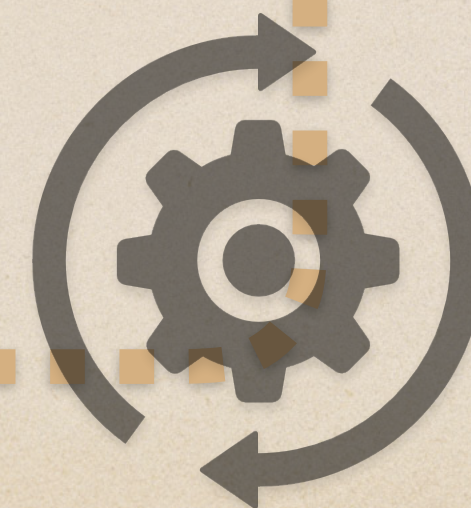## Text Features:

- Title Text
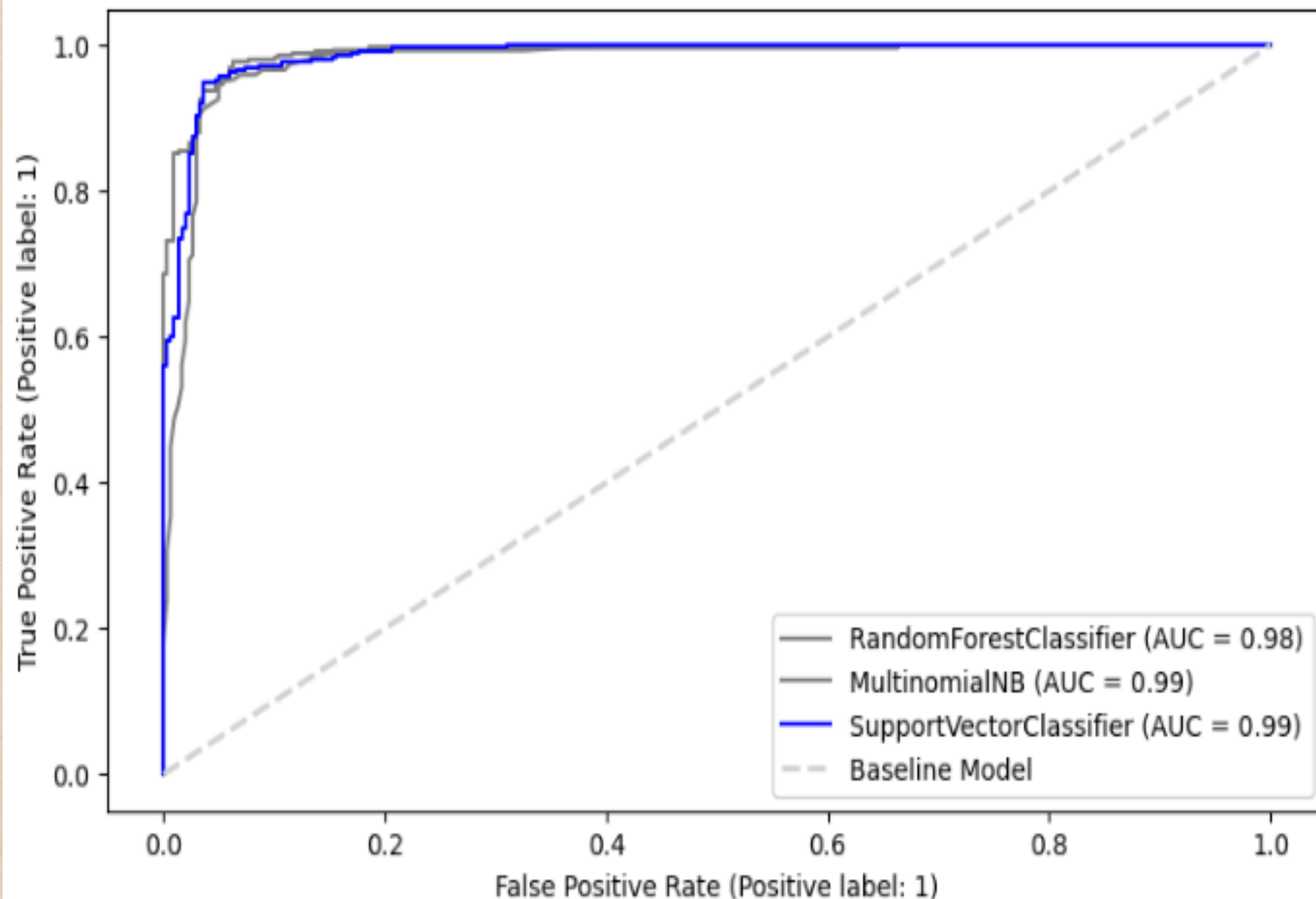- Post Text

MinMaxScalar

OHE

SpaCy Lemm.
TF-IDF Vector.

Classifiers

- Naive Bayes
- Random Forest
- Support Vector Machine

GridSearch Optimisation

Model Evaluation

Accuracy Score

	Default	Optimised
Ran. Forest	93%	94%
Naïve Bayes	87%	94%
SVC	90%	95%

Optimised SVC Confusion Matrix

	Pred: No	Pred: Yes
Act: No	285	15
Act: Yes	14	285

Limitations

Posted by u/somebody12345nobody 1 month ago

Any thoughts on Brandon Sanderson's writing in contrast to Pratchett's?

Discussion

I have a silly little podcast with a friend and he's big into fantasy books. He got me into Lies of Locke Lamora (I'd recommend it up until the third book), The Assassin's Apprentice (not far enough in to give an opinion, and Sanderson's Mistborn Trilogy and Discworld.

Any opinions on how they contrast in world building? Character creation?

Personally, I got so sick of the gratuitous violence, assault, and vague racism that I needed a pallet cleanser and Discworld has been it for me. Mort especially has been a highlight.

Any thoughts? I'd like to hear from someone who has read more Sanderson than I especially, cause I've only read 2 1/2 novels of his.

💬 21 Comments   🎁 Award   ➦ Share   🔖 Save   ···

Bona fide r/discworld posts can sometimes also talk about cosmere

NLP ML models can suggest with high accuracy but will not achieve 100%.

~ GNU PTerry ~