

Natural Language Processing with Reddit

Daniel Preston McBride

DSIR 2-8

19 Mar 2021

65k (PC)

<https://steamcharts.com/app/1085660>

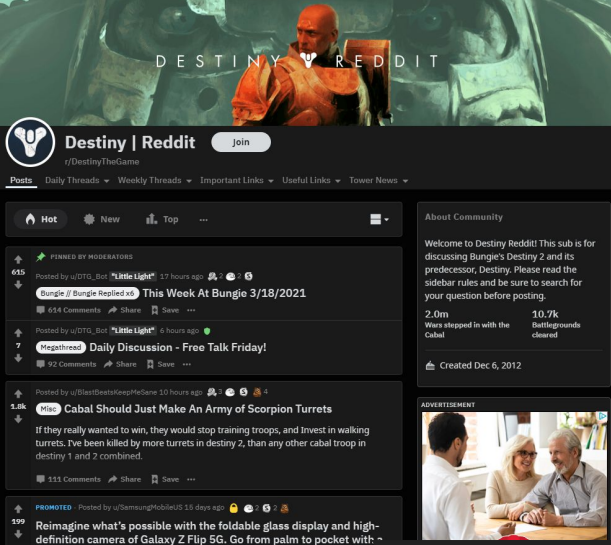
32M

<https://mmo-population.com/r/destiny2>

DESTINY 2

BUNGiE®





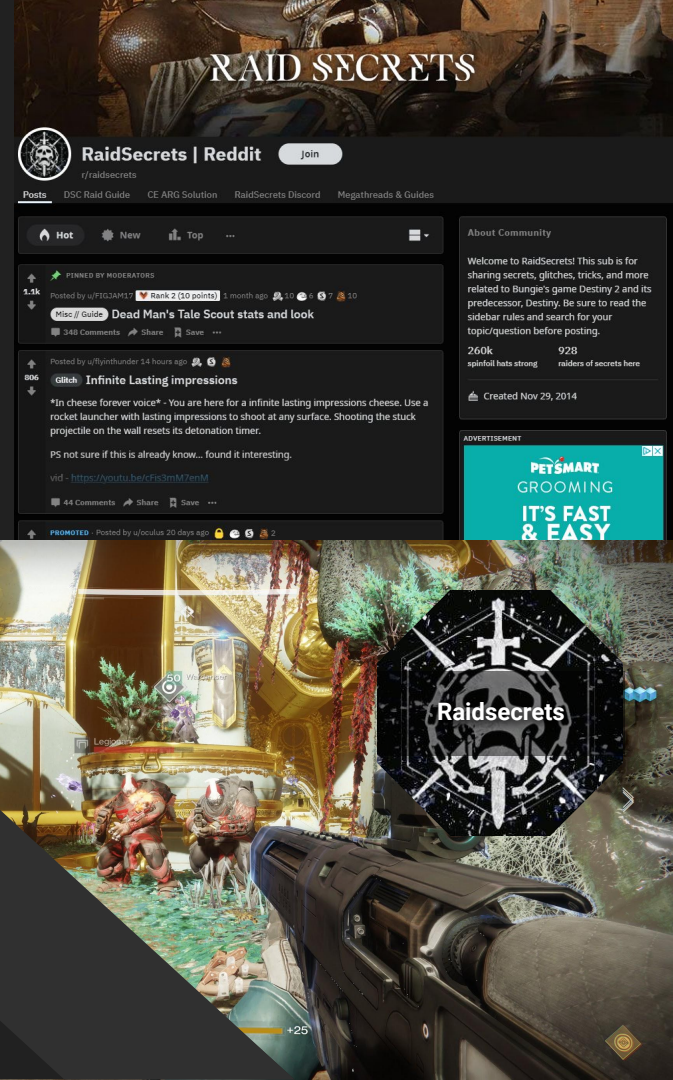
2 million members

- Experiences
- Memes/Jokes
- Feedback
- Ideas



260k members

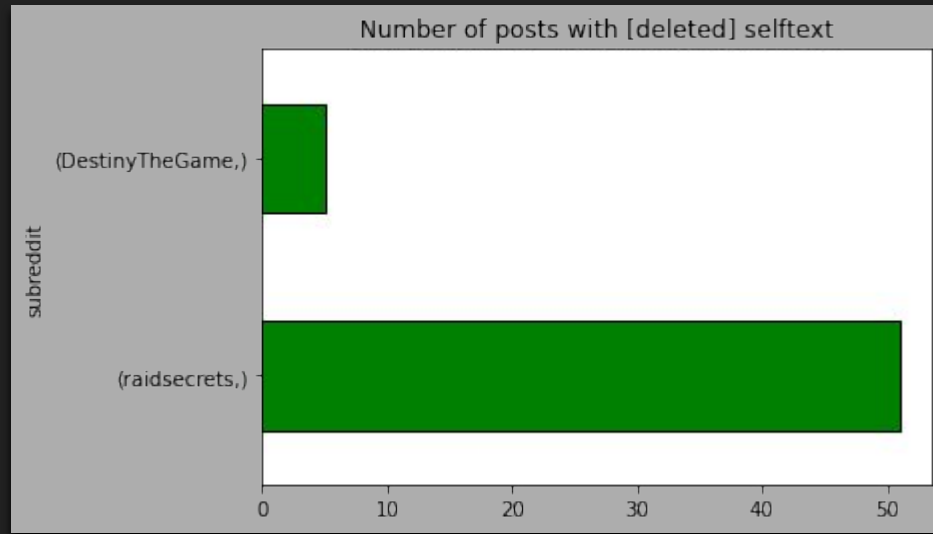
- Secrets
- Glitches
- Tricks
- More...



A man in orange armor stands on a small patch of ground with some plants, looking up at a massive, ancient-looking mechanical skull that dominates the background. The scene is misty and atmospheric. The text "Data Time!" is written in a large, white, serif font, and "Scraping & Cleaning" is written below it in a smaller, white, italicized serif font.

Data Time!

Scraping & Cleaning



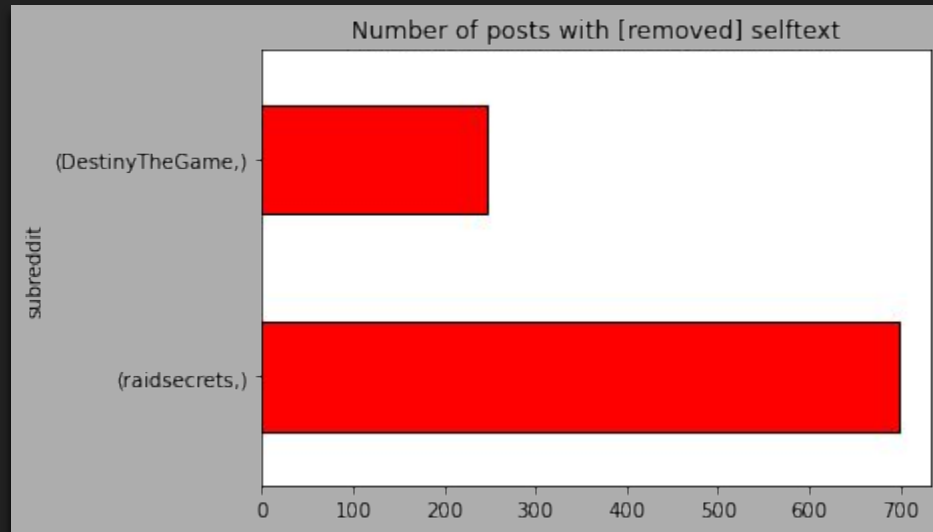
DestinyTheGame



VS



Raidsecrets



DestinyTheGame

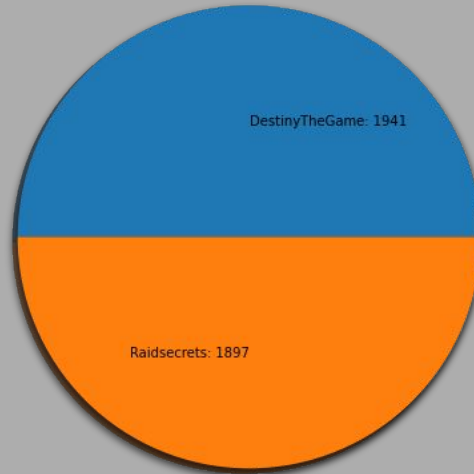


vs



Raidsecrets

Overall Number of Posts Per Subreddit



DestinyTheGame



VS



Raidsecrets

A man in orange armor stands on a small patch of ground, looking up at a massive, ancient mechanical titan. The titan's body is composed of dark, weathered metal with intricate details like gears, rivets, and plates. Its head is a complex structure with multiple eyes and a central mouth-like opening. The scene is set in a misty, overgrown environment with some small plants in the foreground. The overall tone is dramatic and epic.

Modeling Time!

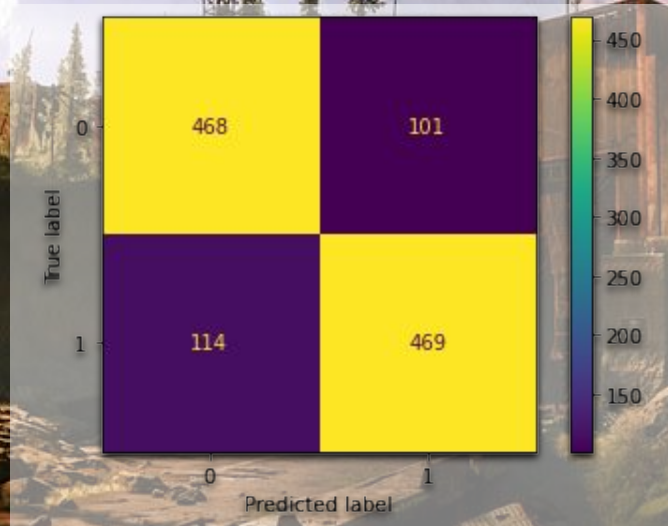
Fitting & Evaluating

CountVectorizer/RandomForestClassifier



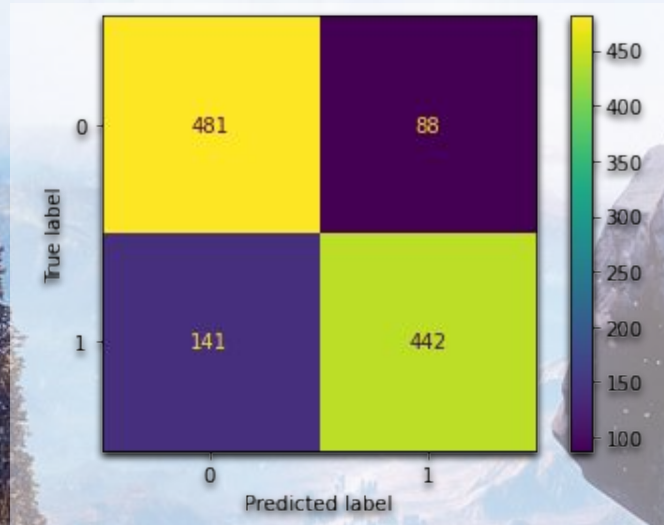
Best Score: 77.2% | Train Score: 78.3% | Test Score: 76.1%

TfidfVectorizer/MultinomialNB



Best Score: 81.7% | Train Score: 85.1% | Test Score: 81.3%

CountVectorizer/MultinomialNB



Best Score: 81.4% | Train Score: 83.9% | Test Score: 80.1%

In Conclusion

Roughly 80% Accuracy across all models



Hope For The Future...

- Gathering more data
- Dropping rows w/only links
- Mod removed posts dropped
- Analyzing most common words

THAT'S ALL FOLKS!!!

Sources:

<https://www.reddit.com/r/DestinyTheGame/>

<https://www.reddit.com/r/raidsecrets/>

