Natural Language Processing (NLP) Machine Learning:

Dungeons and Dragons and Call of Duty

May 29, 2020

Jamie Shaffer

Agenda

Question: What differentiates the player communities for **Call of Duty** (CoD) and **Dungeons and Dragons** (DnD)?

This project explores text analysis as a way of gaining insight into these communities.
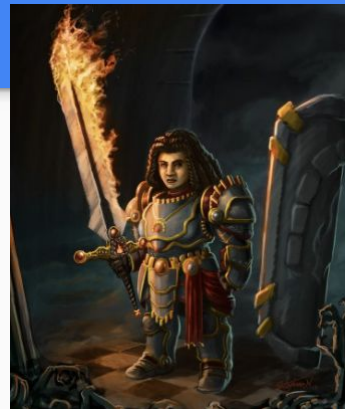
- **Overview of the 2 games**
- **Gather Data**
- **Explore Data**
- **Model**
- **Evaluation**
- **Conclusion**

Overview of the 2 games

Call of Duty

Dungeons and Dragons

- **Low fantasy** - military combat gear
- **Solo play** is the norm
- Co-op play is unusual
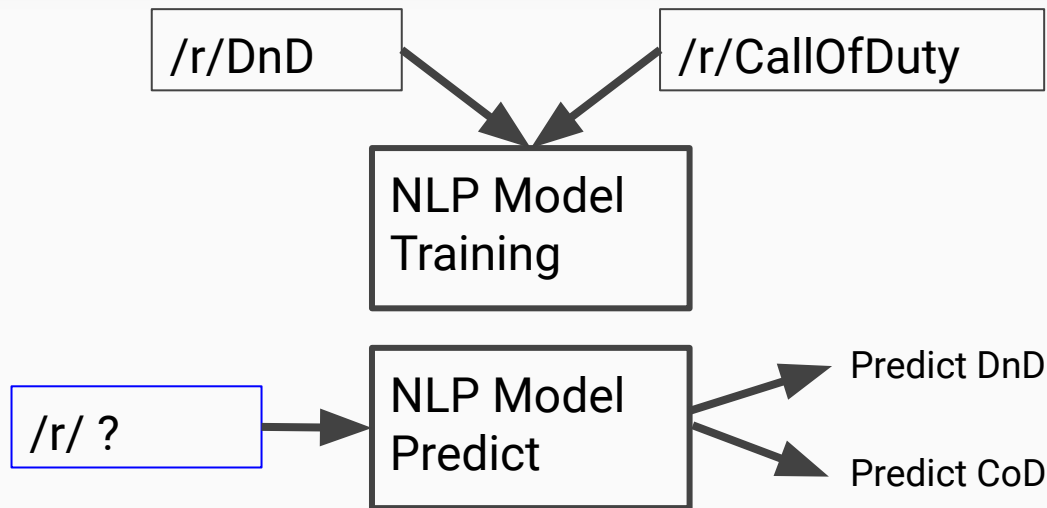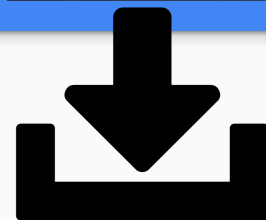- Player vs Player is an option
- **Console** play (XBOX + …)

- **High fantasy** - dragons, magic
- Solo play is unusual
- **Co-op play** is the norm
- Player vs Player is discouraged
- **Table top** (in person + online …)

Question: What differentiates the player communities for CoD and DnD?

1. Using posts from reddit

2. Build a model that can predict which thread the post is from

3. Explore the pieces that have predictive power

```
/r/DnD        /r/CallOfDuty
        ↓      ↓
     NLP Model
      Training

/r/ ? →  NLP Model   → Predict DnD
         Predict     → Predict CoD
```

What might those predictors say about the communities?

Gathering posts from reddit

Using the Pushshift API:

- Programmatically downloaded 500 posts per request
- 8 times
- = 4000 posts from each subreddit

DnD: 9 days of posts   5/20/2020 - 5/28/2020
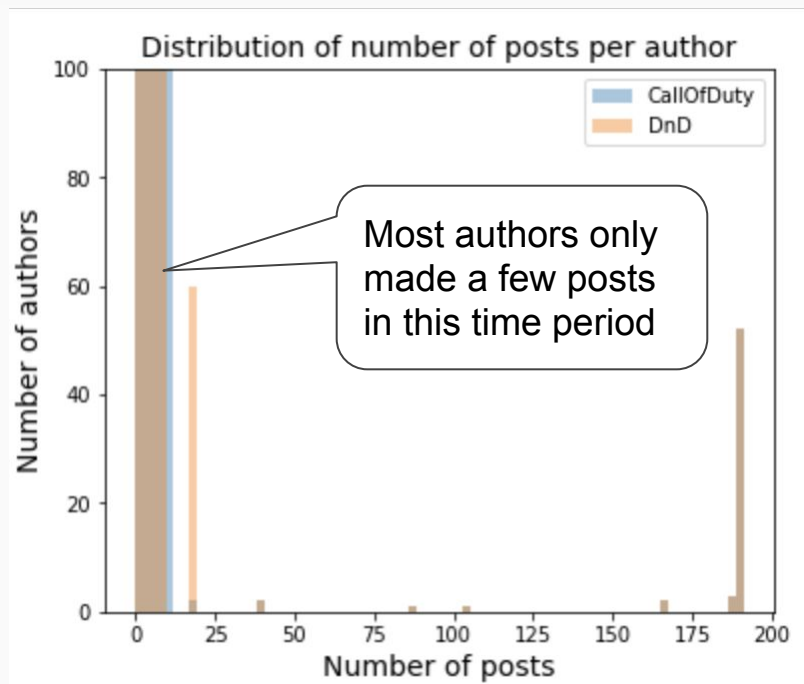
CallOfDuty: 17 (8d more)   5/12/2020 - 5/28/2020

Exploring posts from reddit

Removed 33 banned posts and 166 duplicates

8000 posts → 7801 posts

- 5441 unique authors
- 2 authors who post in both subreddits
- On average:
  - DnD: 1.39 posts per author
  - CallOfDuty: 1.48 posts per author

- Posts: 49.8% CoD, 50.2% DnD

Distribution of number of posts per author

Most authors only made a few posts in this time period

NLP Models

Set aside 25% of the data to test the model

Prepared the text

- Discarded non-predictive words (a, the, of)
- Kept 1, 2, and 3 word phrases

Built 4+ models using the other 75% of the data

- Accuracy of predictions
- Repeatable accuracy on new data

can we please ban this cheater were is the best place for me to post to get him banned thanks
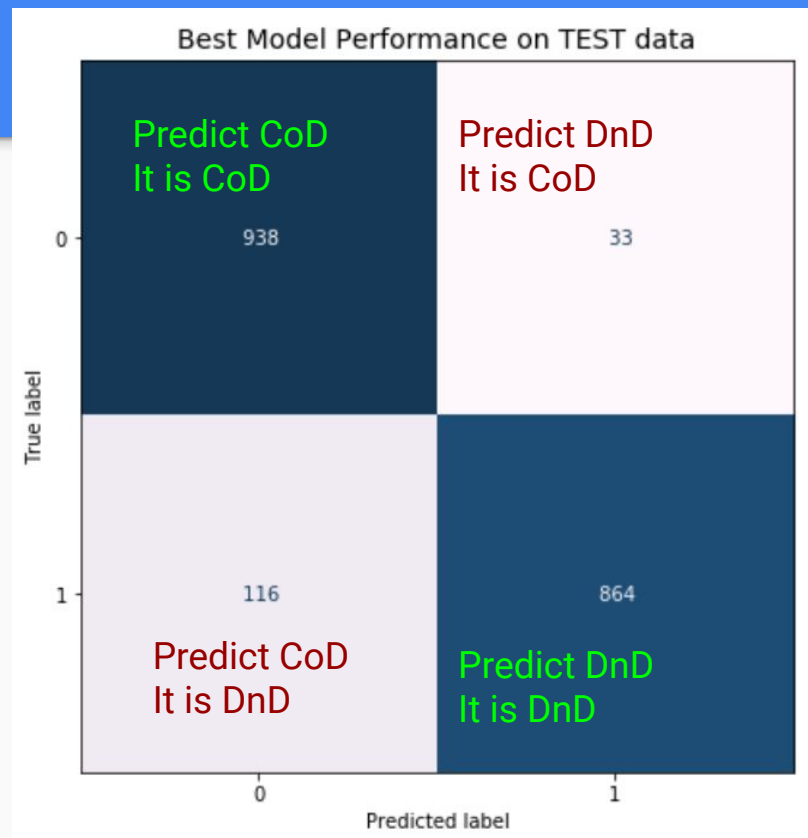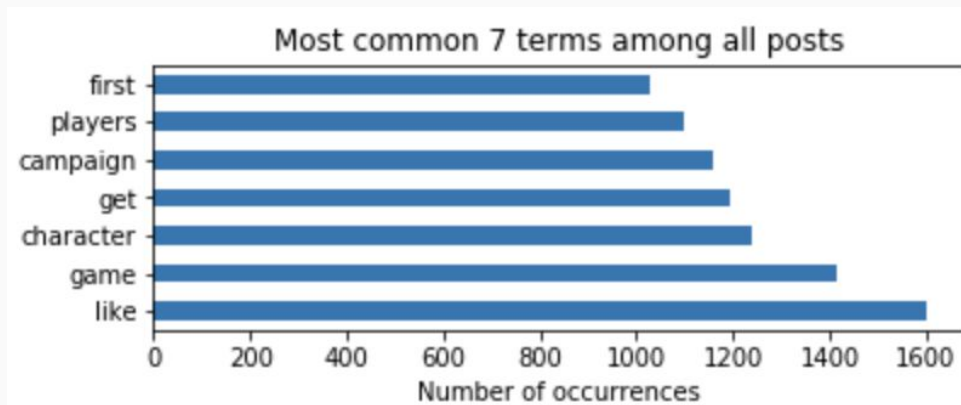
what s going on in your campaign

comment if you don t see your favorite gun this is just what I normally hear

Evaluation:

Baseline of 50% if we guessed 'DnD' every time

Models predict correctly on 92% of posts

(93% if allowed to use 'giveaway' terms like the name of the game)

Most common 7 terms among all posts

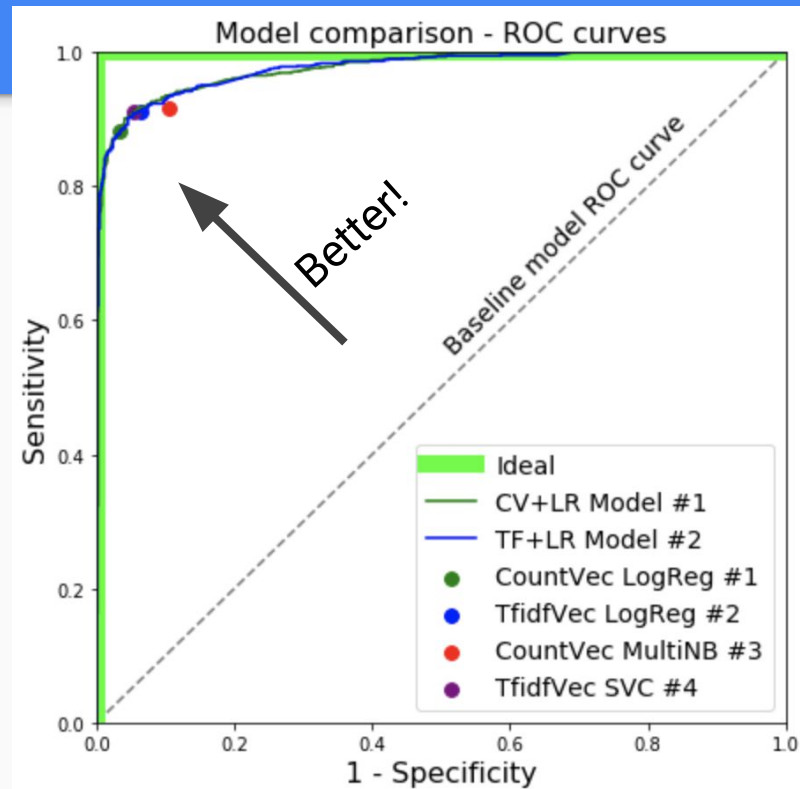

Best Model Performance on TEST data

Conclusion: NLP model can differentiate

NLP (Natural Language Processing) model can correctly predict whether a post is from a CoD or DnD subreddit

- 93% accuracy when including game names (e.g. dnd, 5e, mw, mw2)

- **92% accuracy otherwise**

What might those predictors say about the communities?

Call of Duty

Top Predictors:

Dungeons and Dragons

Call of Duty

Dungeons and Dragons

Top Predictors:

- warzone
- multiplayer
- **remastered**
- lobby
- ww2 / wwii / waw
- mission
- **opinion**
- mobile
- zombies

- oc (original character)
- **art**
- dm
- dice
- bard
- character
- **homebrew**
- monster
- druid

Thank you

Questions?

Addendum

4 Models:

1. CountVectorizer + LogisticRegression
2. TfidfVectorizer + LogisticRegression

3. CountVectorizer + MultinomialNB
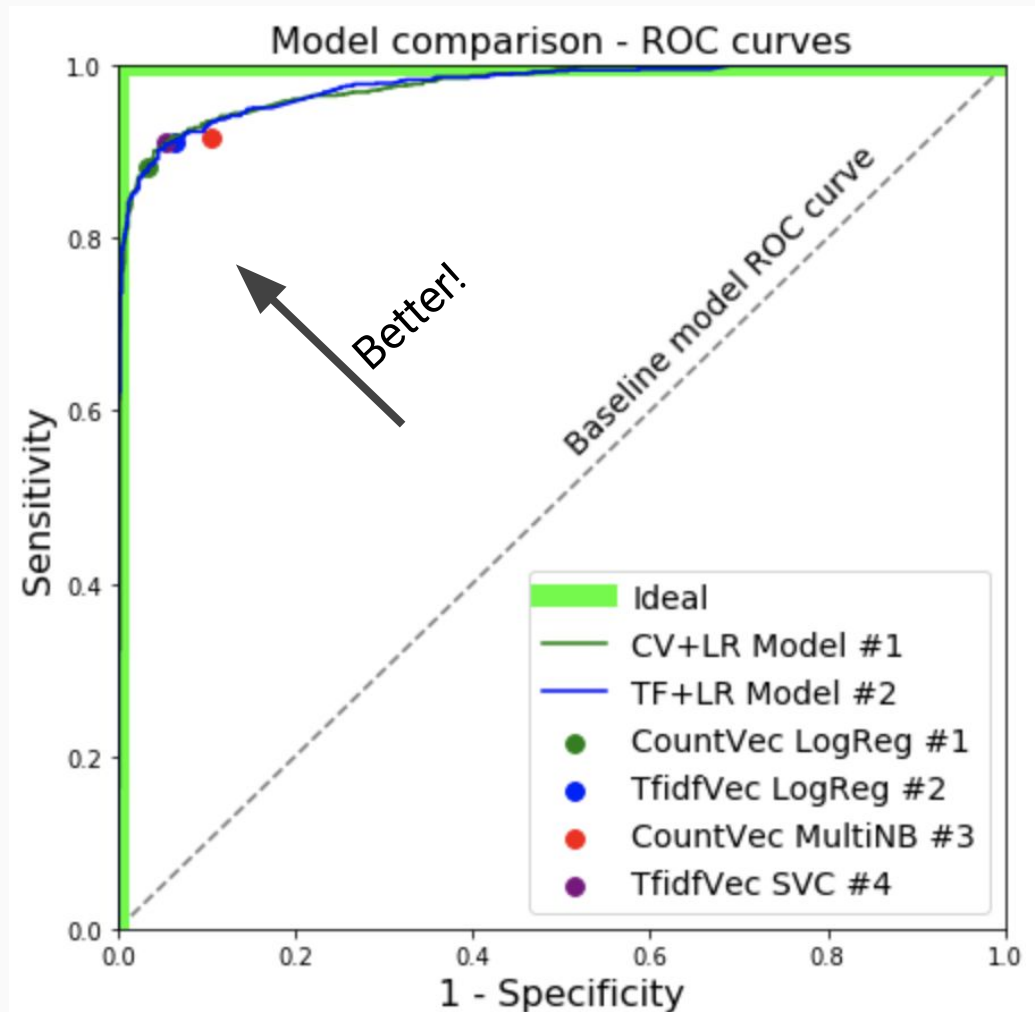4. TfidfVectorizer + SupportVectorMachine

Model comparison - ROC curves

Photo Sources

https://i.ytimg.com/vi/cq5MxnMEiu4/maxresdefault.jpg

https://computerbackgroundimages.com/wp-content/uploads/2018/09/Call-of-Duty-Black-Ops-4-Zombies.jpg

https://i.ytimg.com/vi/M05J-5akzaE/maxresdefault.jpg

https://i.ytimg.com/vi/a_sUqkktG7Y/maxresdefault.jpg

https://basementscientist.files.wordpress.com/2014/04/dungeons_and_dragons_game1.jpg

https://www.ddo.com/sites/default/files/styles/gallery_thumbnail/public/gallery_images/DDO%20Q3%202013%2016-9%20C.jpg?itok=hQjSMl5K

https://i.pinimg.com/736x/52/01/10/5201107008c1de39748e3854612cae4e.jpg

https://i.pinimg.com/originals/da/c6/3b/dac63b04a846a2c491bae4c8e80a44f2.jpg