

YICHEN HU

MODEL TO DIFFERENTIATE BETWEEN 2 POKEMON GO SUBREDDITS

BACKGROUND ON SUBREDDITS

- ▶ r/pokemongo

- ▶ 2 million subscribers

- ▶ Subreddit for discussion regarding all things Pokemon Go

- ▶ r/theSilphRoad

- ▶ 442k subscribers

- ▶ Supposedly focused more on researching the mechanics of the game

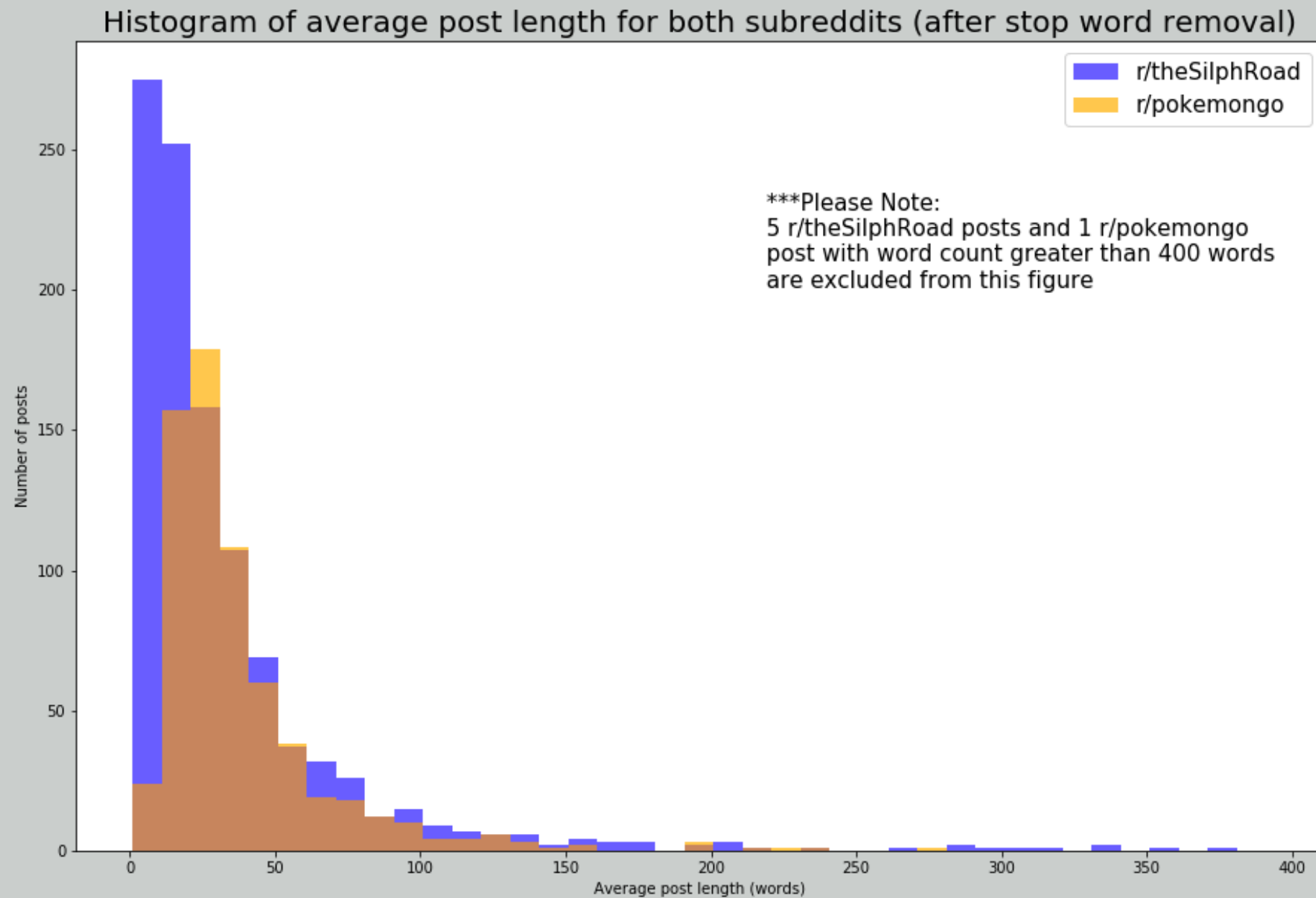
OBJECTIVE

- ▶ Create a model to determine if a post belongs to r/pokemongo or r/theSilphRoad and evaluate it on accuracy
- ▶ Identify if there are certain words/phrases that are used much more on one subreddit than the other
- ▶ Determine if we (as in Niantic, the makers of Pokemon Go) need to designate a more technical employee to monitor the r/theSilphRoad subreddit

OBTAINING AND PROCESSING DATA

- ▶ Data was scraped from the 2 subreddits over a period of 3 days, ended up with roughly 1700 posts
- ▶ Missing values / combining title and post text
- ▶ Natural Language Processing
- ▶ Other processing
 - ▶ Website links
 - ▶ Special characters "é" in Pokémon

AVERAGE POST LENGTH (AFTER NLP)



MODEL RESULTS

- ▶ TF-IDF Vectorizer with Multinomial Naive Bayes gave the best score
- ▶ Accuracy score for best model
 - ▶ .7435
- ▶ Accuracy score for baseline model
 - ▶ .6154
- ▶ 20.8% increase over baseline

WORDS ASSOCIATED WITH EACH SUBREDDIT

- ▶ Top 10 words/phrases associated with r/theSilphRoad

- | | |
|-------------|------------|
| 1. Raid | 6. Shiny |
| 2. Go | 7. Go Fest |
| 3. Dortmund | 8. Mewtwo |
| 4. Pokemon | 9. Spawn |
| 5. Fest | 10. Anyone |

- ▶ Top 10 words/phrases associated with r/pokemongo

- | | |
|-----------------|--------------------|
| 1. Request | 6. Playing Pokemon |
| 2. Ghost | 7. Gameplay |
| 3. Charizard | 8. Finding |
| 4. Caught Shiny | 9. Gyrados |
| 5. Defending | 10. First Shiny |

CONCLUSION

- ▶ I was able to build a model to predict whether a post belonged to r/theSilphRoad or r/pokemongo that was an improvement over the baseline model
- ▶ From the results of my analysis, I would not say the employee(s) assigned to monitor theSilphRoad subreddit need to be more technical