

DC Comics vs. Marvel

Shuya Chen DSI-111



Overview

- ★ Problem Statement
- ★ Data
- ★ Model and Evaluation
- ★ Conclusion





Problem Statement

- ★ Build and train a classifier model to predict the correct subreddit origin of a given post



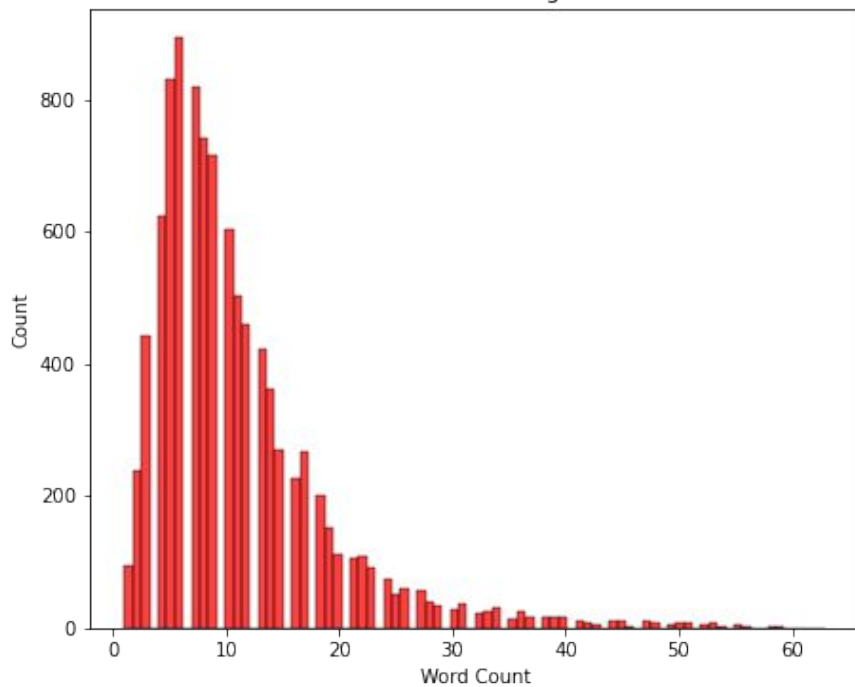
Data

- ★ Pushshift's API
- ★ 10,000 most recent submissions
- ★ 18827 rows, 2 columns

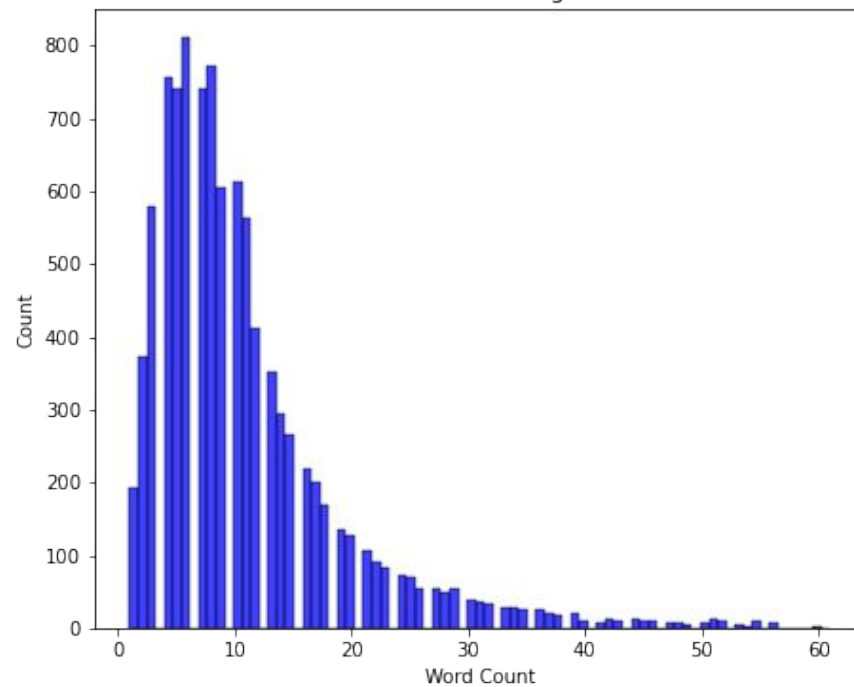


Word Count Distribution

DC Comics

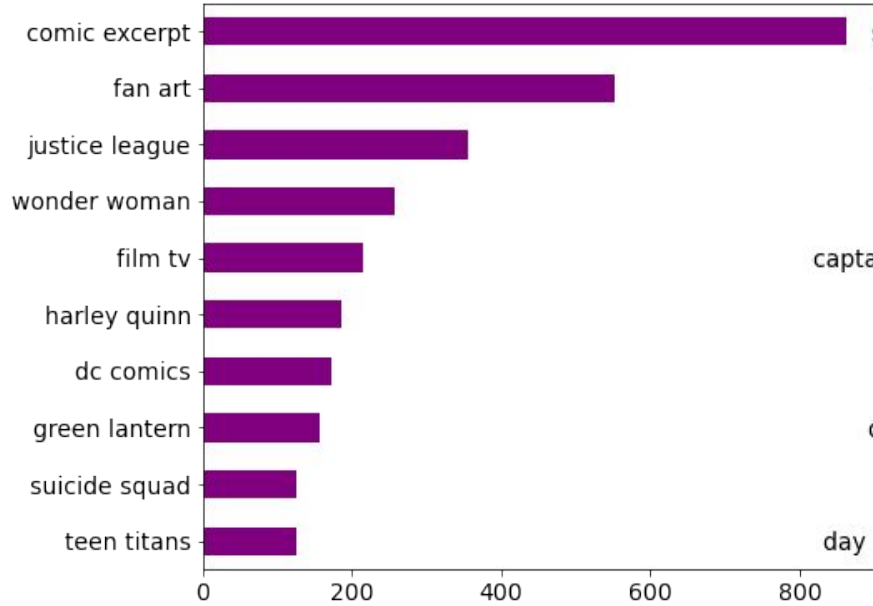


Marvel

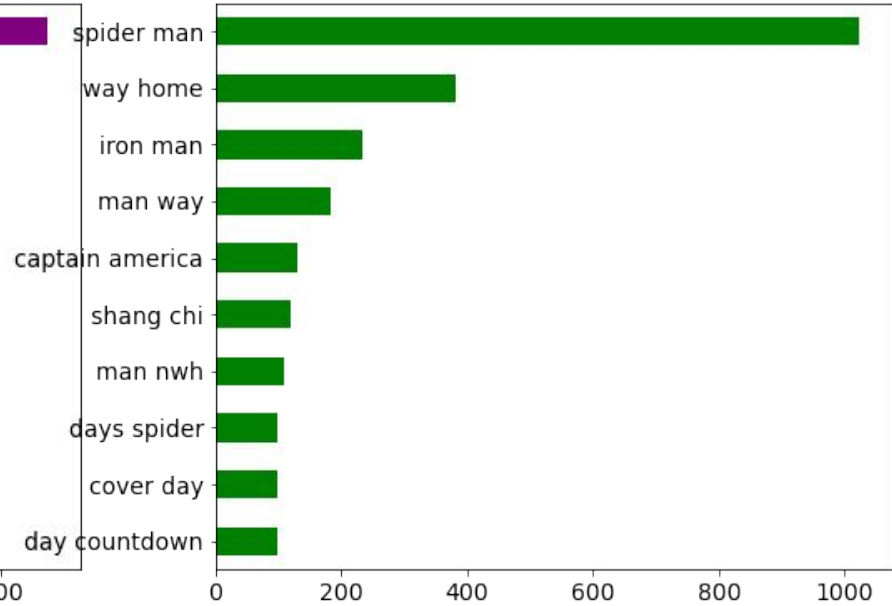


Top 10 Bigrams

DC Comics



Marvel



Number of Occurrences

Model

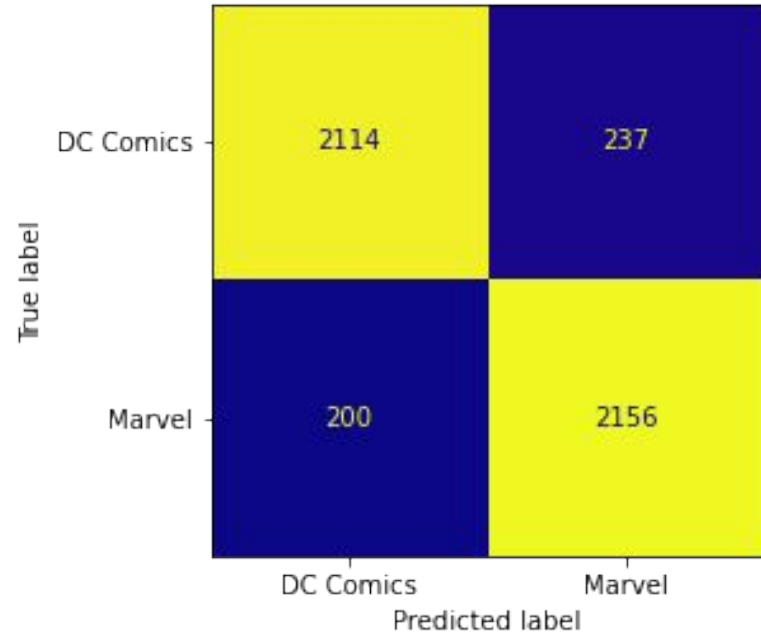
★ Baseline accuracy

1	0.500425
0	0.499575

★ Multinomial Naive Bayes

Train	0.9296
Test	0.9072

★ 90.7 % Accuracy



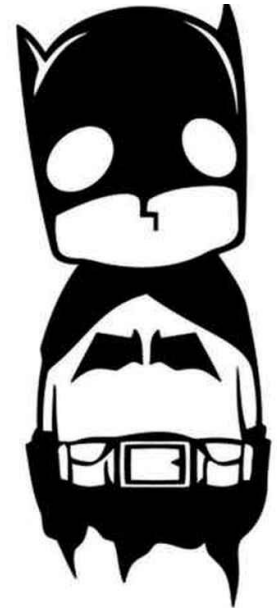
Conclusion

- ★ Good accuracy
- ★ Explore additional models and parameters
- ★ More data
- ★ Comments





Questions



Reference

- ★ https://www.pngitem.com/middle/hhTmhmh_captain-marvel-clipart-ironman-iron-man-cartoon-drawing/
- ★ <https://www.pinterest.com/pin/424464333607003356/>
- ★ <https://www.openculture.com/2015/08/the-1982-dc-comics-style-guide-is-online.html>
- ★ <https://blipfuzz.com/dc-comics-chibi-batman-decal/>
- ★ <https://www.vectorstock.com/royalty-free-vectors/blank-super-hero-speech-bubble-vectors>
- ★ https://www.teahub.io/viewwp/iThbmmx_99-best-superhero-images-on-superhero-bat-family/