APIs

Webscraping

NLP

ZHUOQUAN CHEN

```
ant to be a place where you can ask any and all questions relating to the board gaming world: general or speci
commendations, rule clarifications, definitions of terms/acronyms, and other quick questions that might not war
eir own post. You can see previous versions of this post [here](https://www.reddit.com/r/boardgames/search?
%22%2Fr%2Fboardgames+Daily+Discussion+and+Game+Recommendations%22&amp;restrict_sr=on&amp;sort=new&amp;t=all).
u are seeking game recommendations you will get better responses if you give us enough background to help you.
e [this template](https://www.reddit.com/r/boardgames/wiki/personalized-game-recommendation-template-no-explai
 so. [Here](https://www.reddit.com/r/boardgames/wiki/personalized-game-recommendation-template) is a version wi
planations of what we're looking for.  \n\nHelp people identify your game suggestions easily by bolding the gam
's easy! Just surround the game name with two asterisks (\\*\\*) and it will show up bold. If you reply to any
at has a game name in **bold** with \"**/u/r2d8 getparentinfo**\", one of our robots will tell you more about t
me\n\nJust remember that this is a community full of awesome, helpful people, and feel free to ask your questio
atever they may be. As always, keep it civil and remember what the downvote button is actually for.\n\nLastly
e some of the resources available at /r/boardgames:\n\n* If you are new here, be sure to check out our [Communi
idelines](https://www.reddit.com/r/boardgames/wiki/community)\n* If you are looking for a game recommendation,
re to read our [What Should I Get Wiki](https://www.reddit.com/r/boardgames/wiki/wsig)\n* If asking people for
commendations is too interactive: [The Automated Recommendations Robot is here for you]
ttps://www.reddit.com/r/boardgames/wiki/wsig#wiki_ask_a_robot_for_a_list_of_recommendations)\n* For recommenda
at take accessibility concerns into account, check out [MeepleLikeUs](https://meeplelikeus.co.uk/recommender-be
eir recommender.\n* Also take a look at our [Wiki Index](https://www.reddit.com/r/boardgames/wiki/index) for va
her resources and information.\n* Any questions about game design can be answered here, but also visit
/tabletopgamedesign\n* Any discussions about the subreddit itself should be contained in /r/metaboardgames\n\n
e the boardgamerecommender bot in this thread, please reply to the stickied top comment in order to avoid clutt
e posts.*\n\nHappy gaming!", "author_fullname": "t2_6l4z3", "saved": false, "mod_reason_title": null, "gilded"
licked": false, "title": "/r/boardgames Daily Discussion and Game Recommendations (October 21, 2020)",
ink_flair_richtext": [{"e": "text", "t": "Daily Game Recs"}], "subreddit_name_prefixed": "r/boardgames", "hidde
lse, "pwls": 6, "link_flair_css_class": null, "downs": 0, "thumbnail_height": null, "top_awarded_type": null,
ide_score": false, "name": "t3_jf6lcz", "quarantine": false, "link_flair_text_color": "light", "upvote_ratio":
uthor_flair_background_color": "#d0fffd", "subreddit_type": "public", "ups": 11, "total_awards_received": 0,
edia_embed": {}, "thumbnail_width": null, "author_flair_template_id": "326005ee-9cf2-11e8-8b1f-0e8f9a199476",
s_original_content": false, "user_reports": [], "secure_media": null, "is_reddit_media_domain": false, "is_meta
lse, "category": null, "secure_media_embed": {}, "link_flair_text": "Daily Game Recs", "can_mod_post": false,
, "approved_by": null, "author_premium": true, "thumbnail": "self", "edited": false, "author_flair_css_class":
```

Problem Statement

Scraped a game posts from Reddit and developed a Natural Language Processing model that identify which content of posts belong to board games and which content of posts belong to card games. In addition, it would be better if was able to find more insight with all content.

Data Collection

Using API to collect posts from subreddits

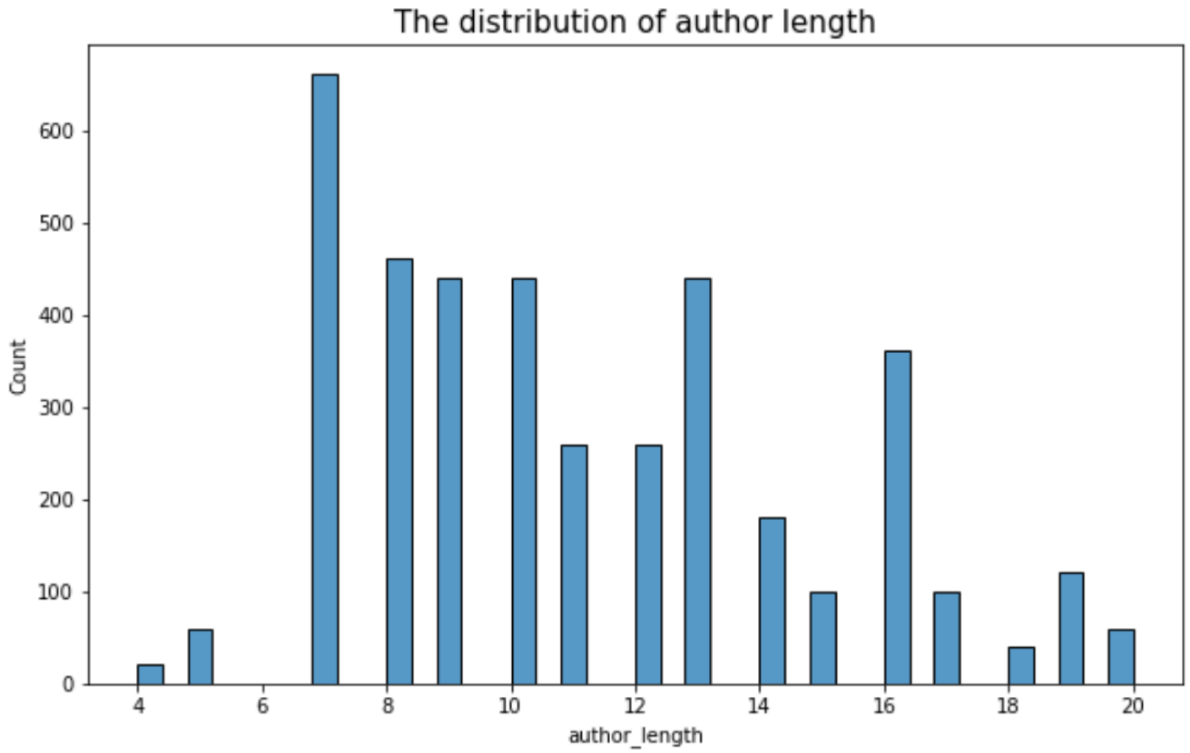
Board Games, Card Games

50 : 50

2000 : 2000

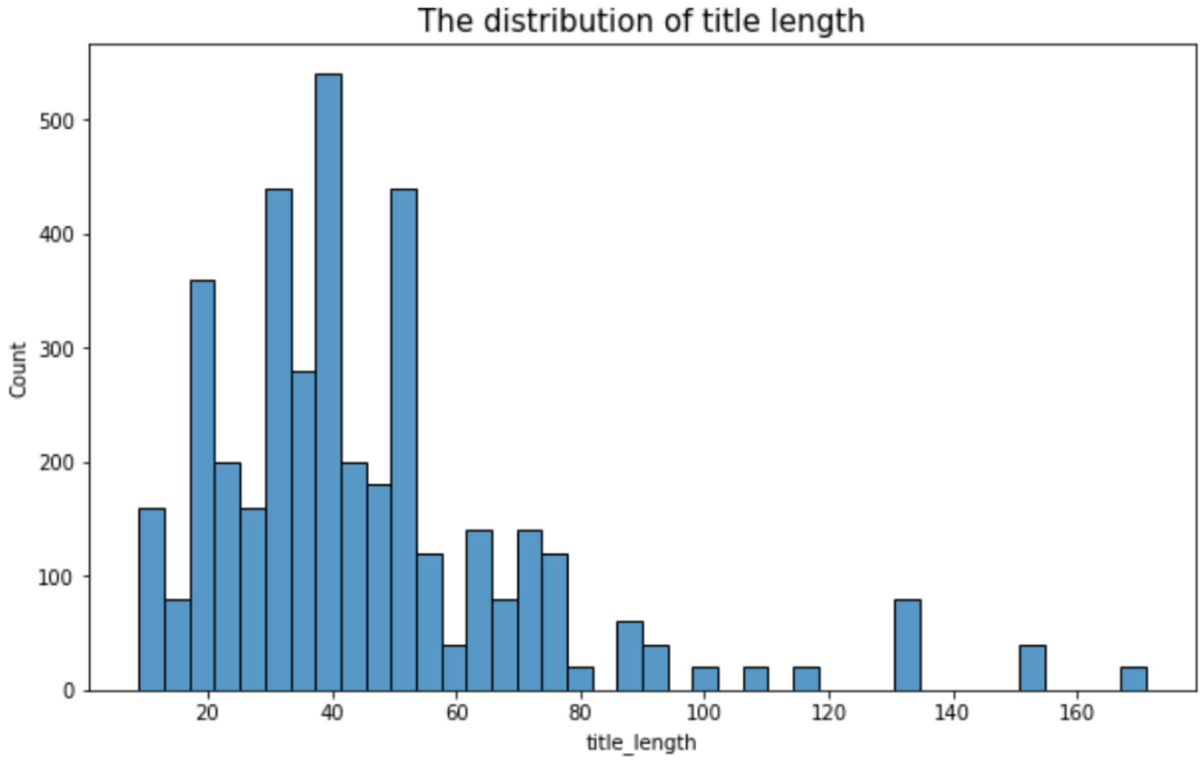
Data Exploration

The distribution of length of authors' names



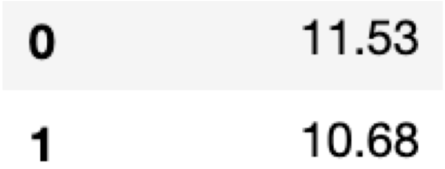
Data Exploration

The distribution of the length of title



Data Exploration

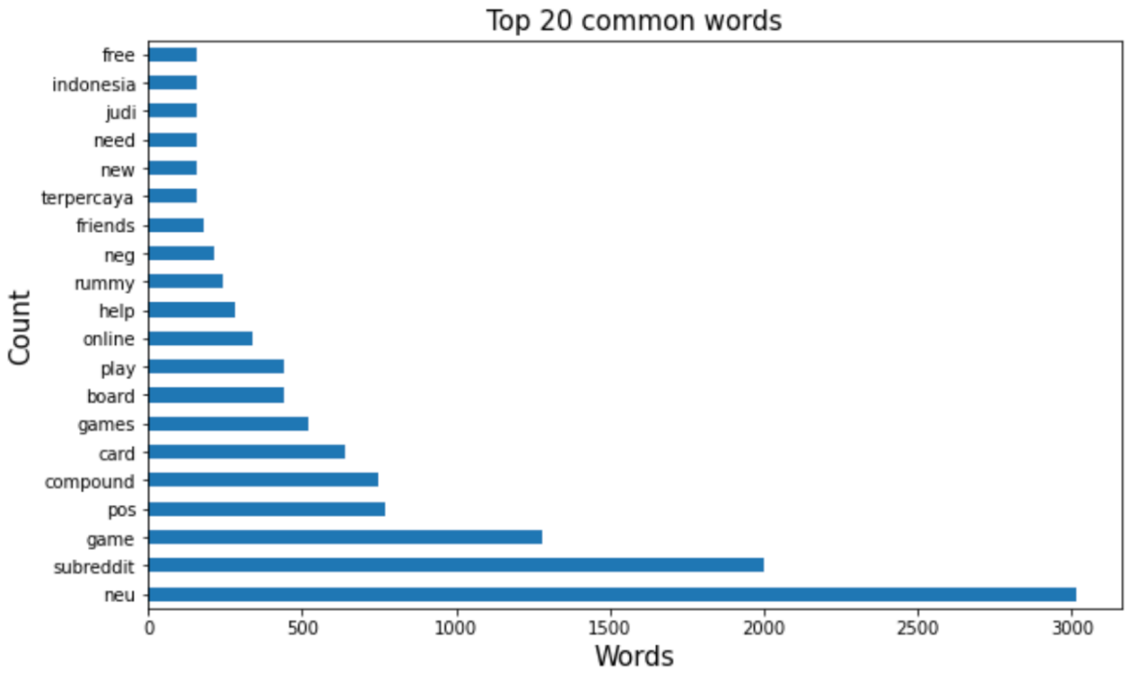
The average of letters of author's names between the classes of board games and card games

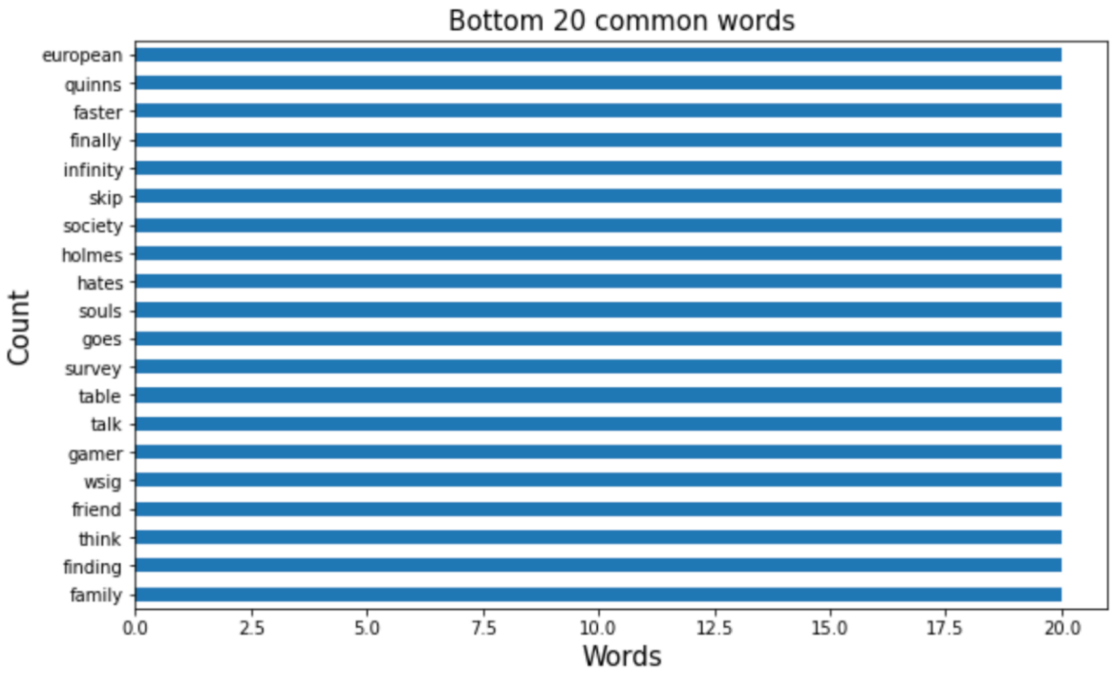


Board Games	0	11.53
Card Games	1	10.68

Data Exploration

The most / least 20 most common words in dataset





Modeling

Logistic Regression

Train Score: **0.637**

Test Score: **0.613**

Hyperparameter tuning & GridSearch

Train Score: **0.622** ↓

Test Score: **0.636** ↑

Modeling

RandomForest

Train Score: **0.866**

Test Score: **0.813**

Hyperparameter tuning & GridSearch

Train Score: **0.866**

Test Score: **0.834** ⬆

Modeling

Gradient Boosting Regression Tree

Train Score: **0.267**

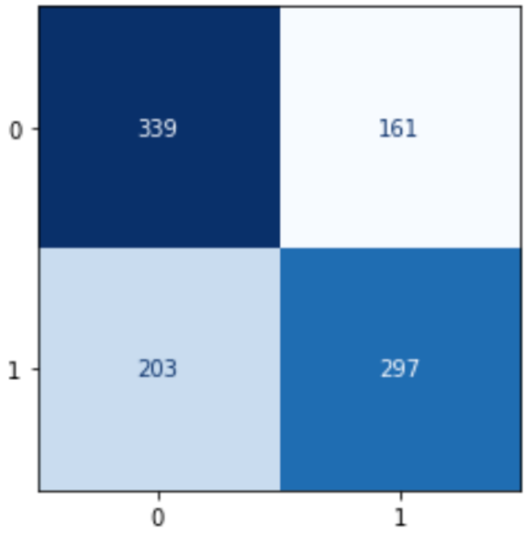
Test Score: **0.260**

Hyperparameter tuning & GridSearch

Train Score: **0.866** ⬆

Test Score: **0.834** ⬆

Conclusion

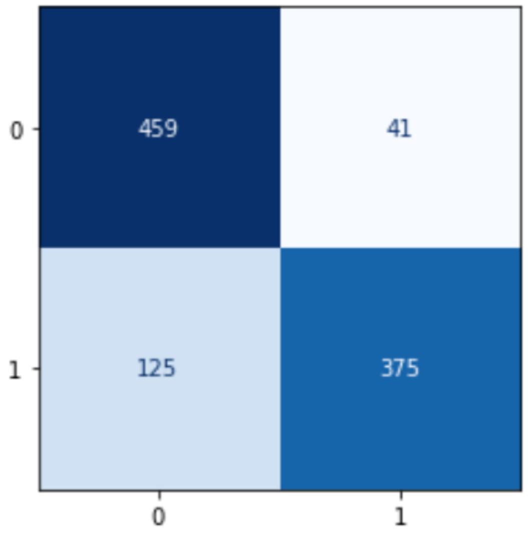


	0	1
0	339	161
1	203	297

Logistic Regression

Sensitivity: 0.625

Specificity: 0.648

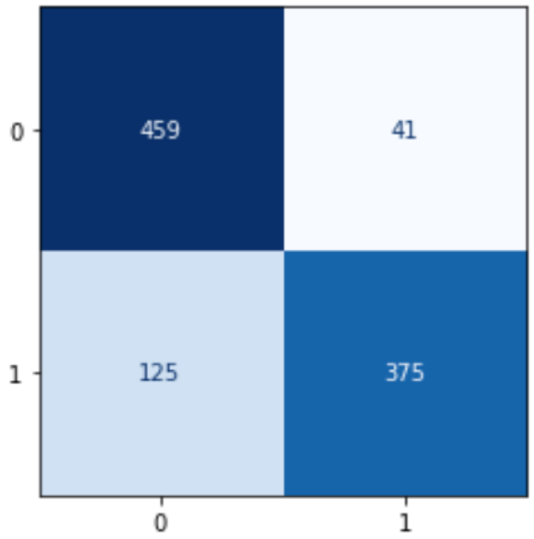


	0	1
0	459	41
1	125	375

RandomForest

Sensitivity: 0.786

Specificity: 0.901



	0	1
0	459	41
1	125	375

GBRT

Sensitivity: 0.785

Specificity: 0.901