

CLASSIFICATION MODELS

Reddit Predictions

Lavanya Bandari

AGENDA

1. Problem Statement
2. Data Gathering
3. NLP
4. Model Iterations
5. Conclusion

Problem Statement



A gaming startup plans to create new boardgames and want to understand what games posts are getting more traction.

Why Boardgames?



[Link](#)

Benefits

1. Board games shown 20% growth in 2020
2. Socially brings family together
3. Easy to buy and affordable
4. Has no age barriers to participate

Understand online behavior

1. Look at reddit posts for board games
2. User behavior patterns for boardgames
3. Any NLP models can predict and give indications on predictions for posts

Data Gathering



(Photo by Sarah Pulliam Bailey)



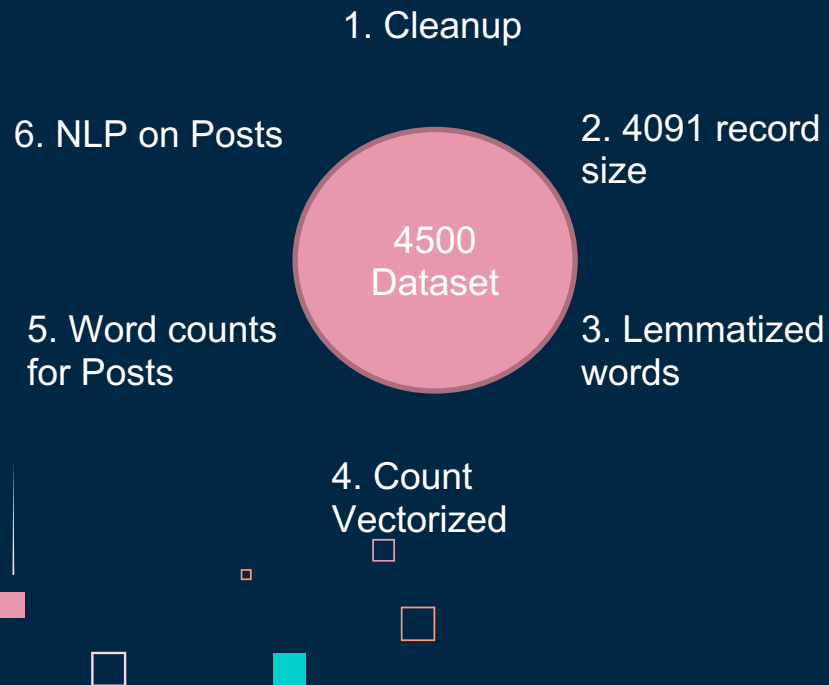
(Photo by pcmag)



(Photo by Bethesda Game Studio)

- Boardgames - 3.4m members , 3.7k Online
- RocketLeague - 956k members, 6.2k Online
- Fallout - 786k members, 3.8k Online

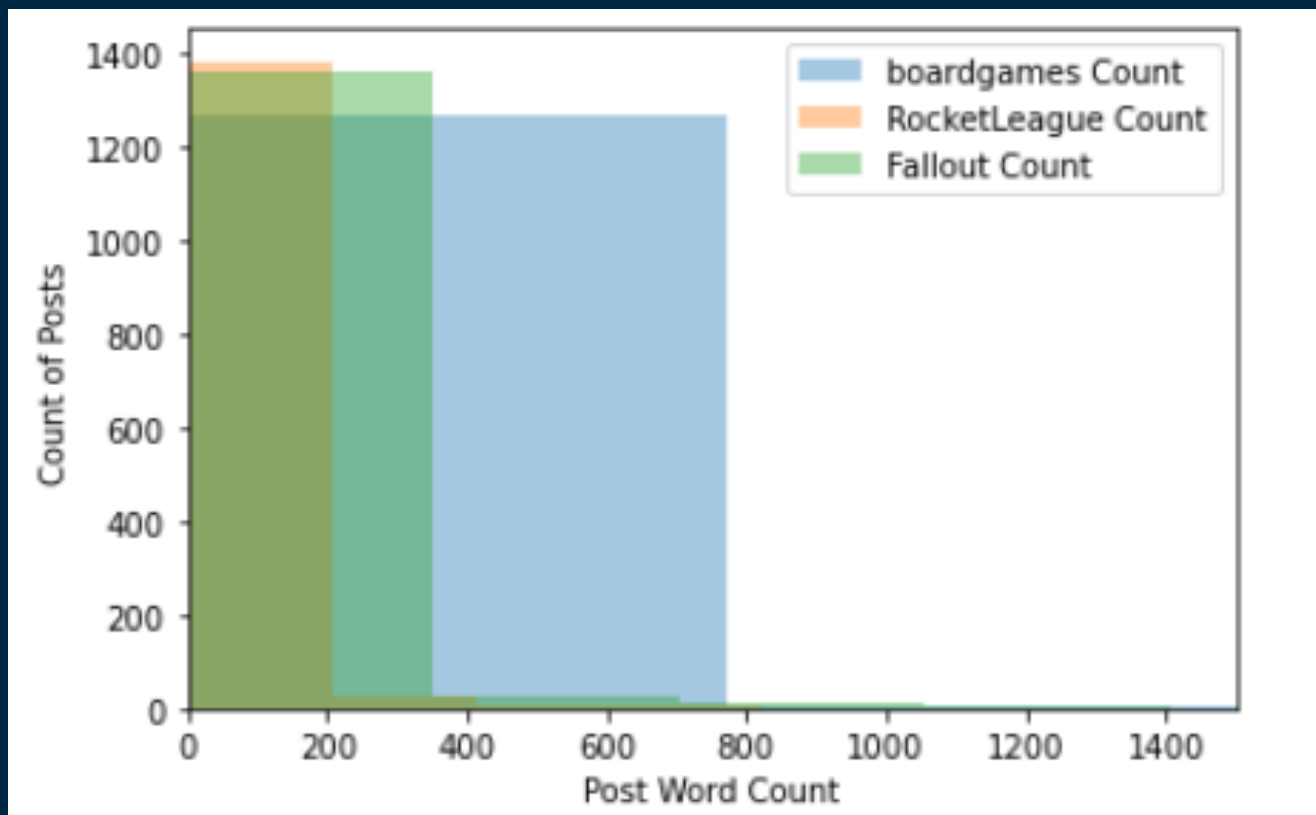
EDA



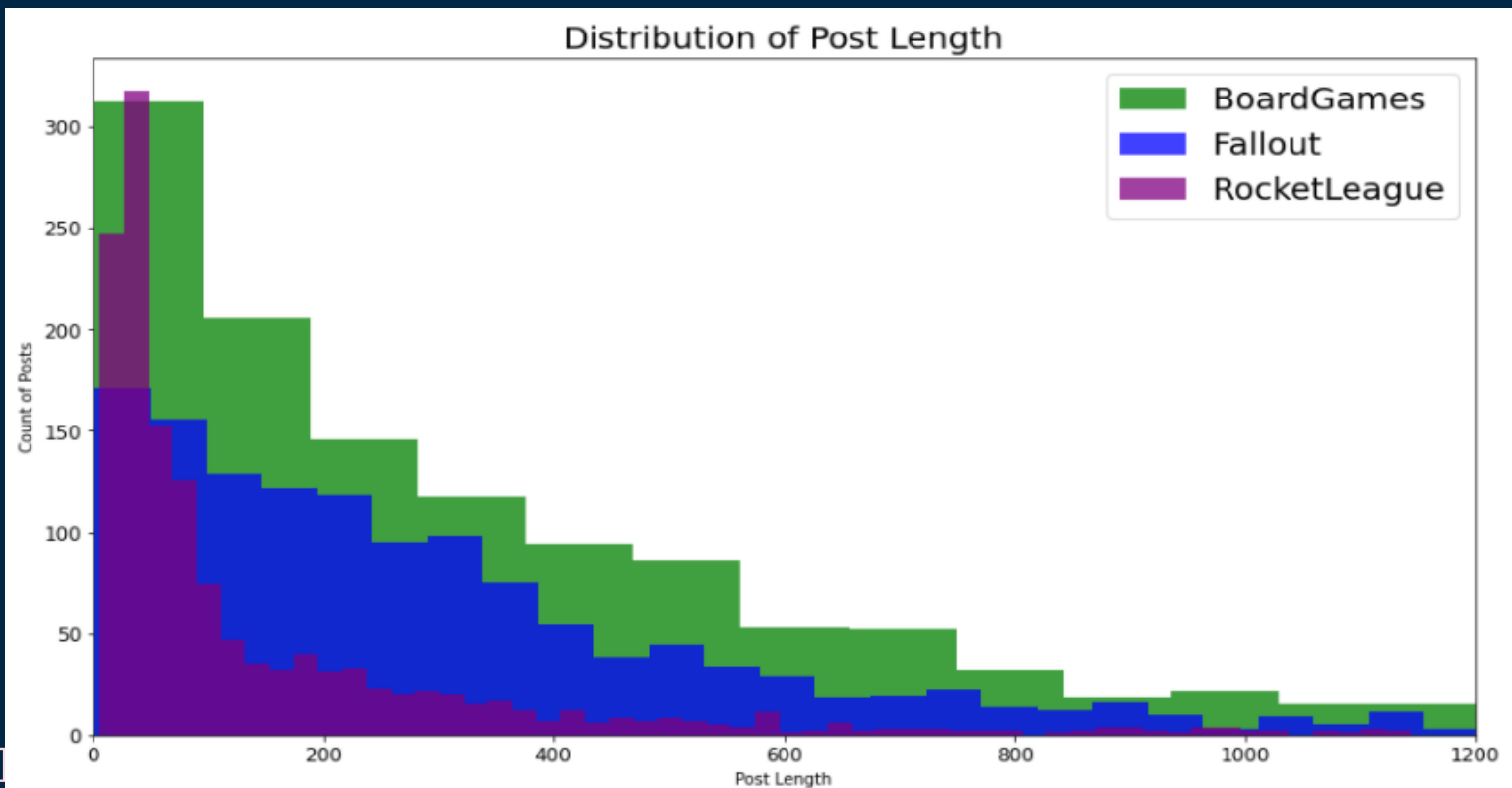
Baseline Normal on classification

Game category	baseline
Fallout	0.342131
RocketLeague	0.346041
boardgames	0.311828

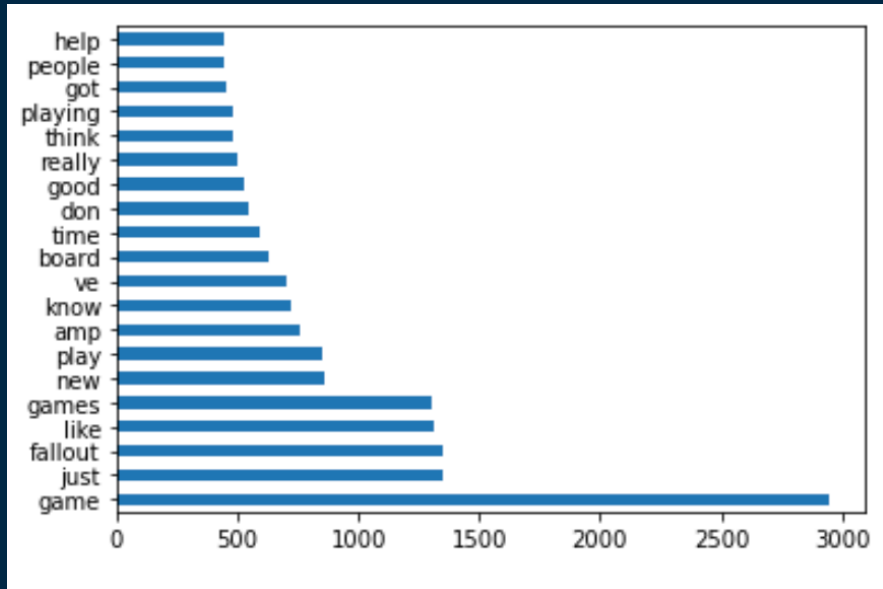
NLP



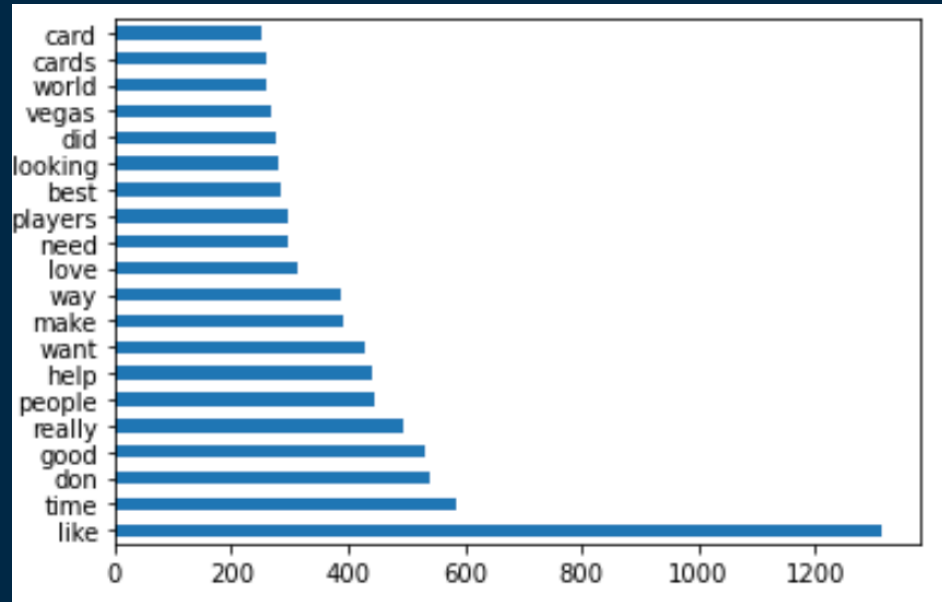
Distribution of Post Length



Stop Words



Before



After

Models

Multinomial Naïve Bayes

- TfidfVectorizer
- GridSearchCV

	precision	recall	f1-score
boardgames	0.896907	0.818182	0.855738
RocketLeague	0.769042	0.884181	0.822602
Fallout	0.781538	0.725714	0.752593
accuracy	0.809384	0.809384	0.809384

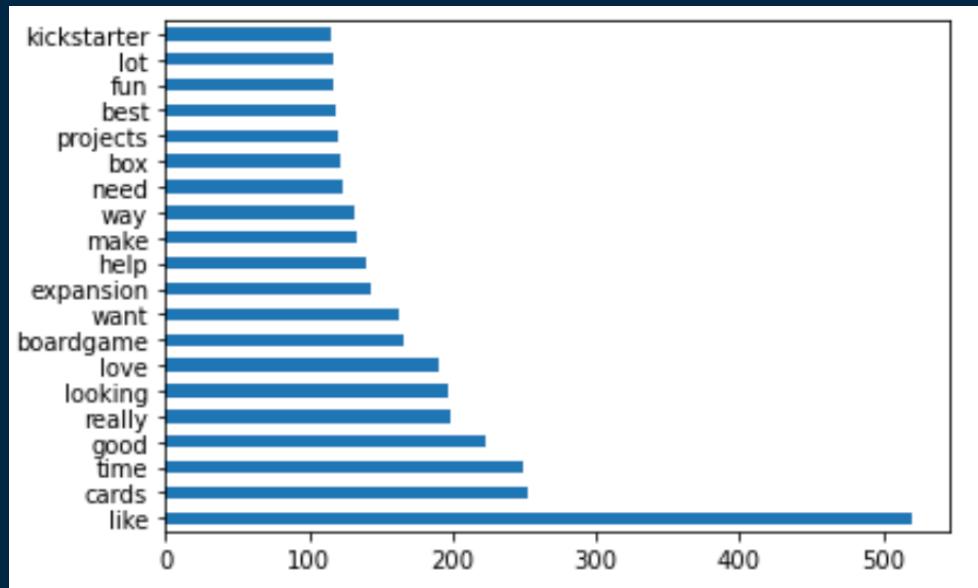
Support Vector Machines – SVC

- CountVectorizer
- GridSearchCV

	precision	recall	f1-score
boardgames	0.899194	0.699060	0.786596
RocketLeague	0.678038	0.898305	0.772783
accuracy	0.757576	0.757576	0.757576
Fallout	0.764706	0.668571	0.713415

Conclusion

- Multinomial NB performed better – slightly overfit
- positive words - like, love, looking, fun, best, kickstarter, expansion, cards, need, help, good, time
- Post word counts did show lot of activity towards Boardgames



The background is a dark blue field decorated with a pattern of small squares and thin vertical lines. The squares are in three colors: light blue, pink, and orange. Some squares are solid, while others are hollow. The vertical lines are thin and white, extending from the top or bottom of the frame. The word "Questions?" is centered in a large, white, sans-serif font.

Questions?