



# Dominion: NLP

Andrew Falcone





# Overview

Can the title of a Reddit post predict how many comments that post will receive in relationship to the median?







# The subreddit: Dominion



# Dominion

- 01 | Deck building card game designed by Donald X. Vaccarino
- 02 | Score more than your opponents by building a better deck
- 03 | Hundreds of available cards with 14 expansions
- 04 | Nearly every game unique with 10 random kingdom cards
- 05 | Reddit: strategy, questions, events, fan cards





## Example Posts

Back after long hiatus

Fan Card – Chasm

What the most fun unique combo you've done?

Missile and Missile Launcher

Fan cards – Missile and Missile Launcher

Fan Card – Blessing in Disguise

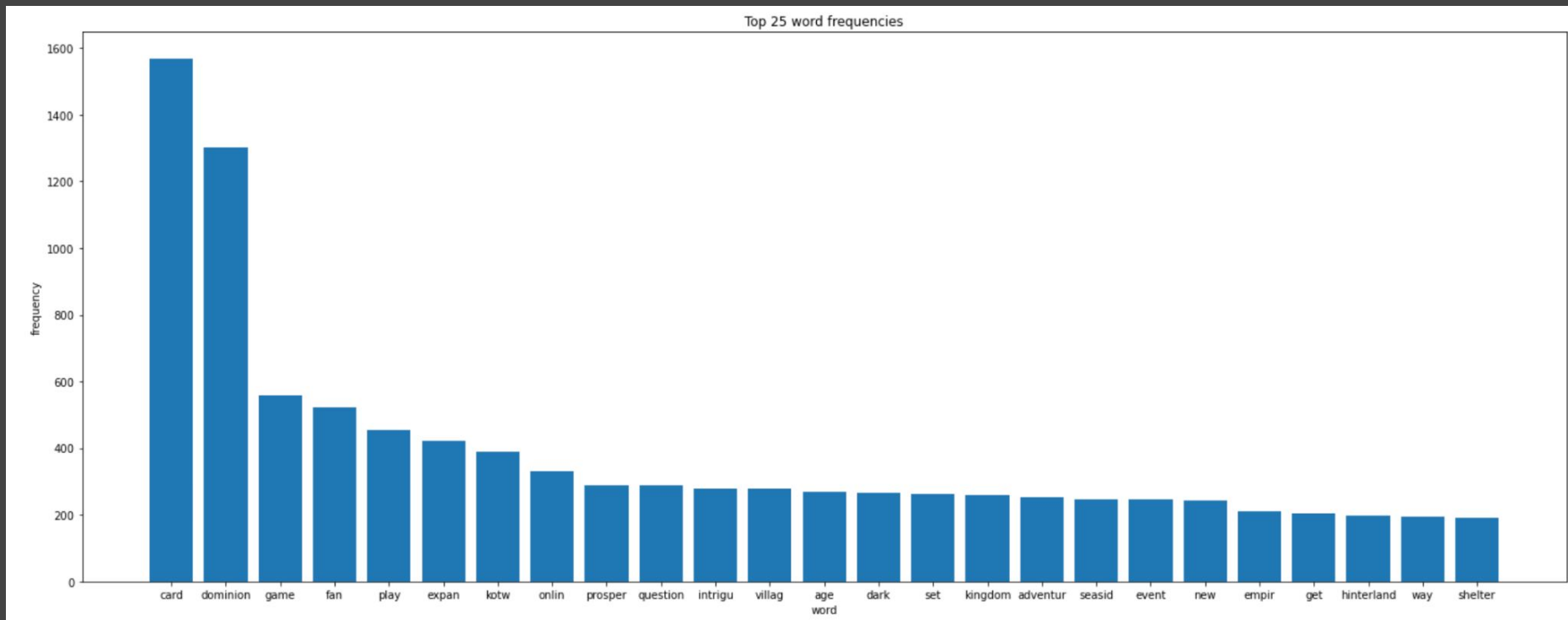
Highway is definitely an experience

Favourite anti combo

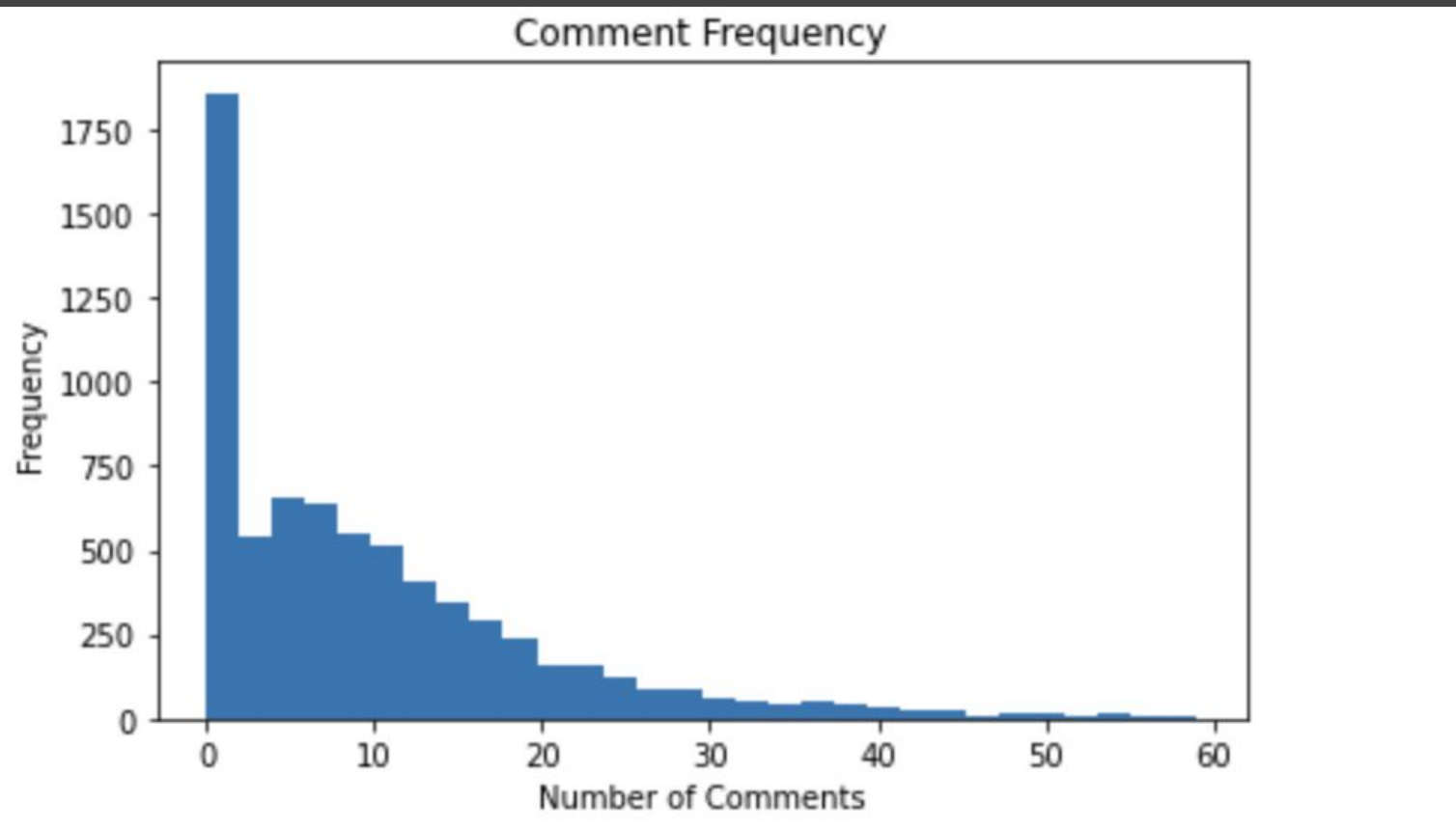
KotW 9/4: Baker, Butcher, Cobbler, Fool, Guide, Herald, Peasant, Ranger,  
Shepherd, Stonemason. Events: Lost Arts, Save. [Guilds, Adventures, Nocturne]



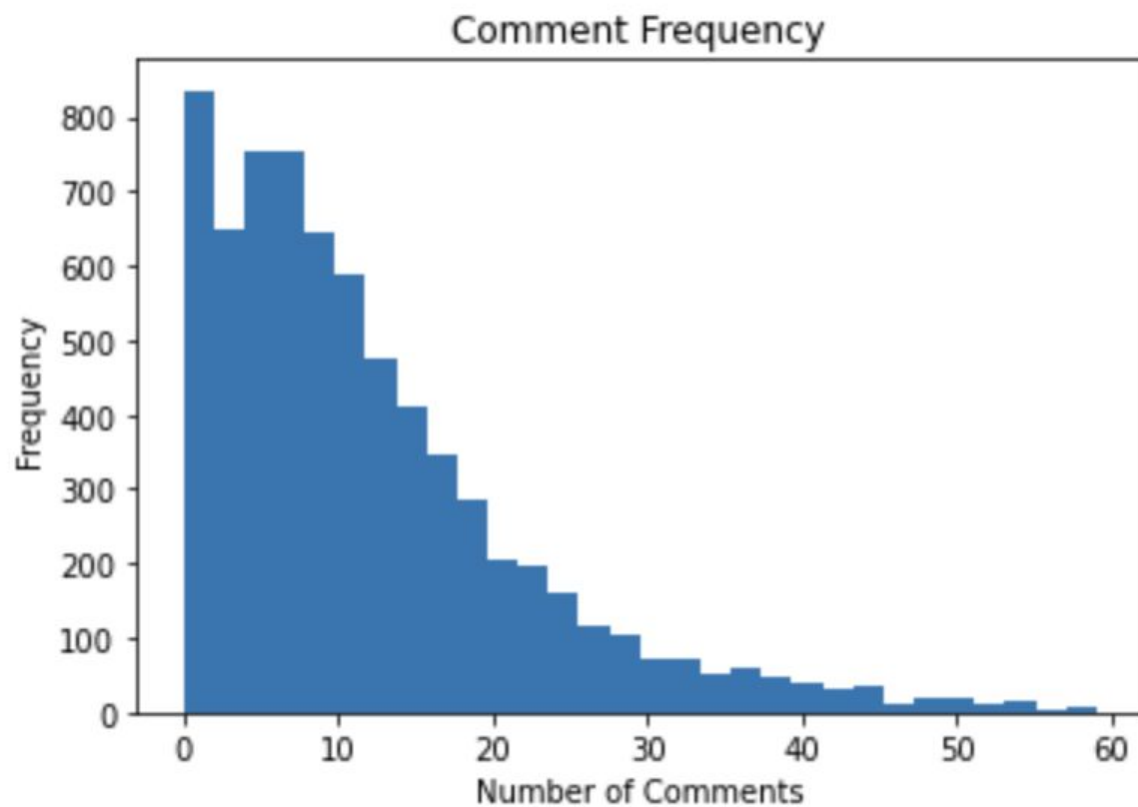
## Most Popular Words



Original

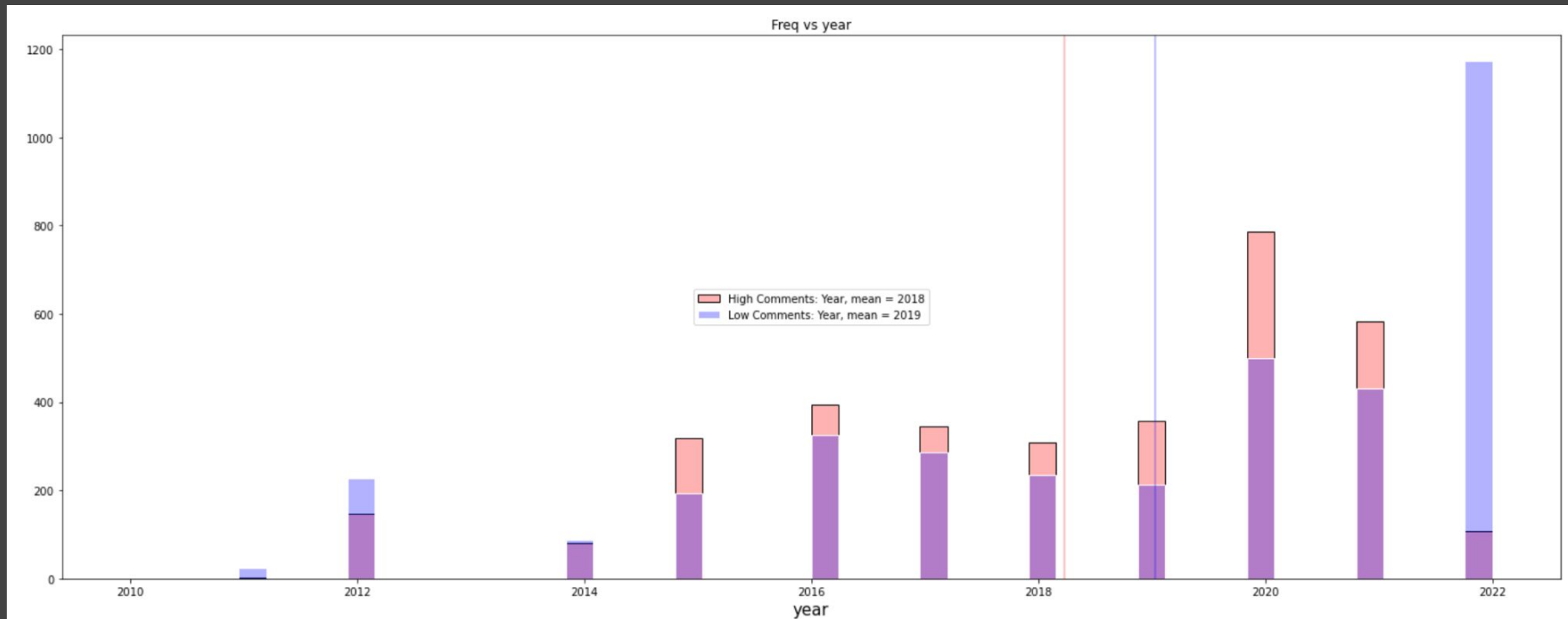


FIXED





## Comments by year





# Models Used

1. Logistic Regression
2. Naïve Bayes (Bernoulli and Multinomial)
3. Random Forest
4. Gradient Boost
5. Ada Boost
6. KNN
7. Extra Trees
8. Support Vector Classifier





And the winner is:

## Support Vector Classifier

Accuracy = 61.7 %  
(Baseline = 51%)



## Posts Predicted to Go Viral

Laminating cards  
Expansion Recommendations based on liked Base Game Cards  
Mini-expansion based on my alma mater (RPI)  
Discard clarification for Renaissance Cards  
Some less-than-half-baked fan card ideas  
Which expansion to buy next for ESL family?  
Should I go simple or more complex with my fan card?  
Opinions on the removed first edition Intrigue cards  
Menagerie Previews 5: More Cards  
Do you take the placeholder cards?





# Posts Predicted not to go Viral

Dominion Storage solution: 5 boxes compacted to 2.  
Voyage bug?  
Dominion League S13 Championship Match – Tomorrow at 18:00 UTC  
Dominion online down?  
Small Games Tournament 3  
Dominion | Recommended Sets  
Small Games Tournament 4  
Online rating  
Way of the Frog Question  
Dominion League S15 Championship Match – Sunday at 18:00 UTC



## Conclusions

- Language is challenging
- Hard to predict engagement
- Model tends to correctly predict posts with unrelated text/spam and titles with 'question' as low virality
- Model has more difficulty with False Negatives than False positives (misses popular posts)



## Recommendations

- Engaged user base makes predictions more difficult than if half data was spam/deleted and low engagement
- Modify goal, this would have a very low human score as well
- Spend more time on grid searching parameters and more rigorously analyze misclassified posts
- Use more features to look for trends in data



# Questions??

# Thank you.

