



# NLP Classification Models and Reddit



Derya Gumustel



# Project goals

---

- Train and compare two predictive models that can identify which subreddit a post came from
- Requires text-rich subreddits, not images and videos
- Initial ideas: oceanography, politics → rejected
- Final choices: r/DMAcademy, r/truezelda, r/PoliticalDiscussion

↑  
251  
↓

Posted by u/AreTooDeeTo 10 hours ago

**Need Advice** How do I handle a player who prepared too much?

This requires a little bit of context I think. Gonna do the best I can to summarize leaving a book of text. Minor TOA spoilers ahead. Running TOA with some mixed in. My group had multiple backstory plot line hooks in Jahaka and were going there to fake as slave buyers, buy one of my characters mother's slavery, and then allow my other player to renounce his pirate background and port to the ground. He has been working as a double agent, feeding the and Liara have given him back into Port Nyanzaru, working against his old friend turned pirate captain.

They get to the port and everything is basically empty. I've laid a trap with character from the module who was with the party for a few sessions but was kicked out for being shady. She went and betrayed the group to the pirates as the group was coming. My group splits themselves, one player to talk with

74 Comments Award Share Save ...

↑  
77  
↓

Posted by u/Shandrih 4 hours ago

**Need Advice** DMing with absolutely zero visual imagination?

I have aphantasia. For those who don't know about it, it means I have no visual imagination. If I asked you to picture a table with a ball on it, you'd probably answer some simple questions, right? What color is the ball, what shape is that sort of thing. I can't. I don't see the ball, so when you ask I have to describe the spot. Or, you know when someone tells you to close your eyes and visualize and you can mentally see it? Or the "movie" you see as you read? I don't see at all. When I close my eyes, the only thing I see is the inside of my eyelids.

So, now that we've got that out of the way, I'm curious to know if there are any aphantasic DMs in here? Any advice for someone who is relatively new to DMing and particularly appreciate some advice on how to keep track of things in combat?

125 Comments Award Share Save ...

↑  
2.2k  
↓

Posted by u/ChattanoogaGuy 1 day ago

**Resource** Worlds Without Number changed my DMing game

I've been playing 5e for a couple years here and there, and last year I decided to lead a group of friends. I took up the mantle because I wanted DnD, and no one else had the gumption to do it. So after having only played about 5 actual sessions, I decided to first and started a homebrew world. I watched a lot of CR, MCDM videos, players manual and DM's guide cover to cover, and scoured the internet for things to look out for, etc. I learned a TON, and fast. I was a sponge!

I adopted the Lazy DM's method, and would only plan what what right in the

↑  
379  
↓

Posted by u/dosfosforos 1 day ago

**Open Discussion** Zelda doesn't work as a multiplayer experience

Well, maybe it has worked in past titles, but it scares me when I see BotW 2 to be a 2 player (or worse, online multiplayer) game. I really believe it's a good idea. I know this could be just me, but I feel like Zelda is a game series out there (excluding FS and TH), you just get in Links' shoes as the hero of Hyrule inside this little giant world in which you make your way to depend on a partner to play through would ruin that experience probably would need to make it possible to play alone as well, which badly designed puzzles or making one of the players essentially useless. Imagine the bad controls (like Bowser Jr in Bowser's Fury), an awkward AI for those who choose single player. And don't even get me started just a bunch of Links running around and lagging like hell on the Nintendo Switch. Nintendo offers, that would suck! The prospect of this happening kills me. I really don't want this to happen. What do you guys think? Would it

97 Comments Award Share Save ...

↑  
11  
↓

Posted by u/Brynmaer 23 hours ago

**Open Discussion** Why BOTW 2 could be a spiritual successor to the Past

With the trailer and a few words in interviews we have a fairly clear idea of what BOTW 2 will have a "darker" tone. Because of the darker tone and all the comparisons have been made with Majora's Mask. I believe Nintendo may find more inspiration for BOTW 2 from a different game in the Past.

We know that BOTW was partially designed as a spiritual return to the roots of Zelda. It was designed with freedom, a large emphasis on exploration and limited direction which allowed the player to discover the world and in almost any order they want.

There is a heavy implication that BOTW 2 will have more of a focus on the story, centered around Ganondorf. The gameplay tradeoff with a focus on the story is a heavy implication that BOTW 2 will have more of a focus on the story, centered around Ganondorf. The gameplay tradeoff with a focus on the story is a heavy implication that BOTW 2 will have more of a focus on the story, centered around Ganondorf.

8 Comments Award Share Save ...

↑  
4  
↓

Posted by u/darklordofftech 19 hours ago

**Game Design/Gameplay** Did you have trouble figuring out how to defeat the Blind The Thief boss fight in A Link To The Past?

Did you walk into the boss room without the "maiden" and/or without the "maiden" and find yourself wondering where the boss is?

5 Comments Award Share Save ...

↑  
26  
↓

thek826 2 days ago

Did you read the article you linked?

Black, Indigenous and other people of color, also known as BIPOC

Reply Give Award Share Report Save

↑  
30  
↓

SpiffShientz 2 days ago

They definitely didn't read it, or they'd know it's a private group funding a small number of families and not a "tangible law"

Reply Give Award Share Report Save

magus678 2 days ago

It not being a law would not stop an outcry were it to specifically exclude black people

4 Reply Give Award Share Report Save

11 more replies

↑  
25  
↓

Shut-up-and-party 2 days ago

Nah, they just wanted to rant. It's ironic that in a time where Asian people are dealing with a wave of discrimination that I'm hearing a lot of Asian men ranting against other people of color.

The original questions to posed by OP is rife with fodder for feeding racist rhetoric anyway, so I would be surprised if there wasn't any of this kind of BS in the thread. It's funny to me as a BIPOC to pose white identity politics as anything other than your run of the mill everyday racism.

1. Is there any room left in this country for racism?
2. If no, how do we end racism?
3. If yes, how do we perpetuate racism?

Hmm. Academic trolling much?

Reply Give Award Share Report Save

↑  
11  
↓

MessiSahib 2 days ago · edited 2 days ago

Nah, they just wanted to rant. It's ironic that in a time where Asian people are dealing with a wave of discrimination that I'm hearing a lot of Asian men ranting against other people of color.

So you want Asians to complain about only some form of discrimination (which can be attributed to right) but ignore other forms of discrimination (which can be attributed to the left).

Reply Give Award Share Report Save

4 more replies

# Text cleaning

- Dropped rows that contained removed, deleted, and null posts
- Used regular expression to remove hyperlinks, digits, and punctuation
- Also removed odd strings like '&amp;'

	text	title	removed_by_category
47	[removed]	Who is the Queen of Air and Darkness? What hap...	moderator
89	[removed]	A helpful video about how to design a balanced...	moderator
224	[removed]	Character backstory assist.	moderator
227	[removed]	New player to Dnd/ build help pls	moderator
344	[removed]	My friend told me if I had a pet that the B...	

Examples of Set-pieces: [Shark vs Rowboat](<https://www.youtube.com/watch?v=c-xQej688fI>),  
[Goblin Bridge](<https://www.youtube.com/watch?v=r9rhPAS3r1U>), [Zombie Shark and Drow Escape](<https://www.youtube.com/watch?v=goyXZOPW1qI>)

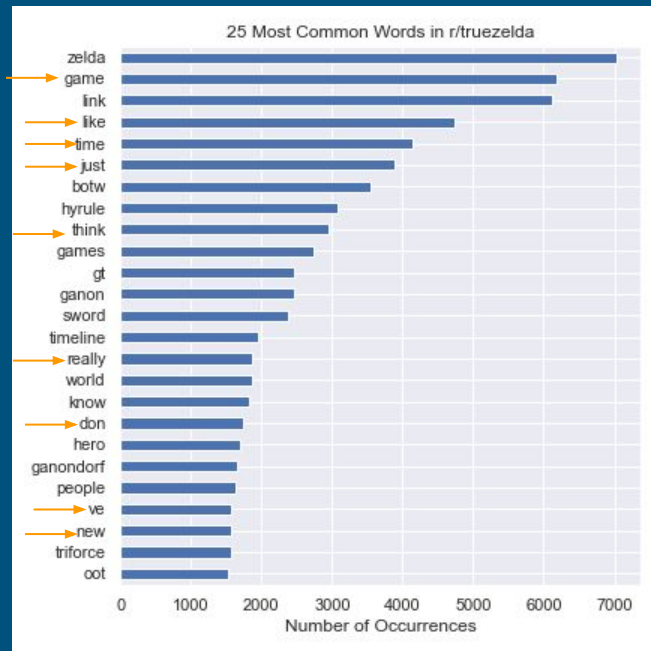
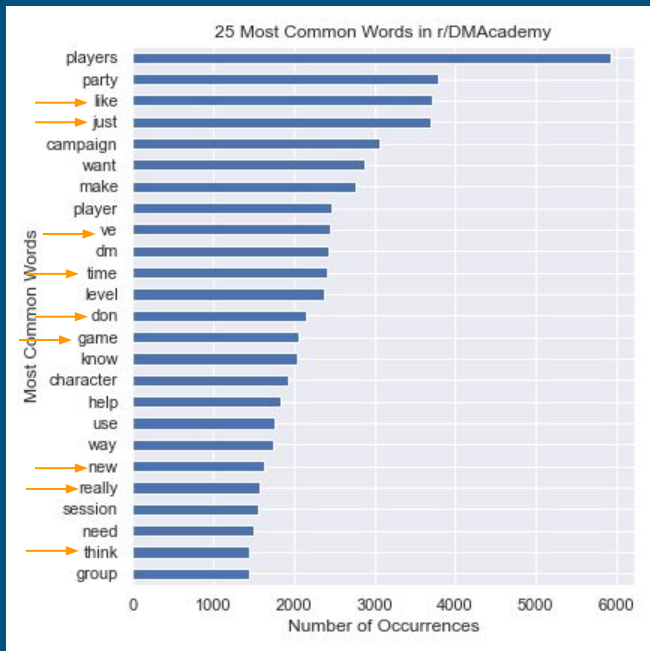
# Text cleaning

- Identified two forms of bot messages
- There are probably others!
- Removed these posts
- Final dataset size: 4887 posts from r/DMAcademy, 4112 posts from r/truezelda

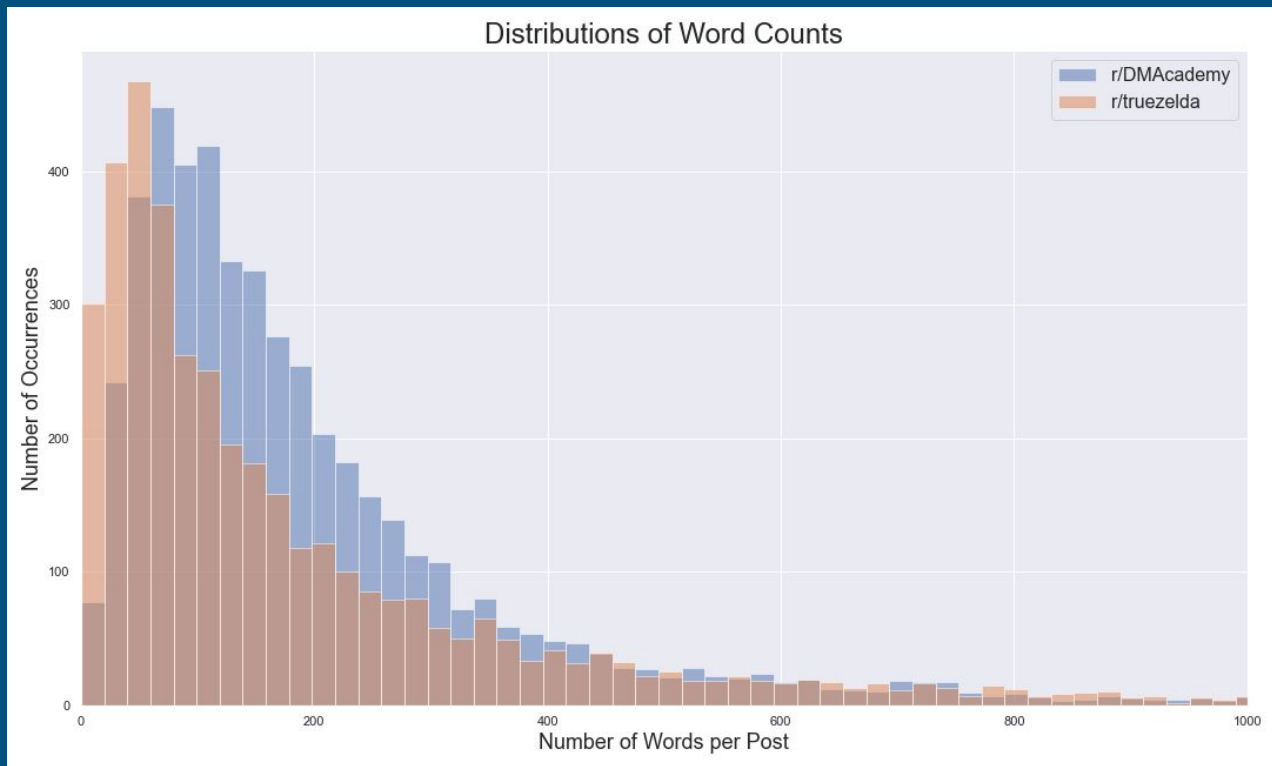
```
poli_dis_2020_df
37      Hello, /u/PMmeURsprintPROGRAMS. Thanks for con...
38      Hello, /u/pleasedontbullyme_. Thanks for contr...
39      Hello, /u/chickenman86. Thanks for contributin...
40      Hello, /u/dumbirds. Thanks for contributing! U...
67      Hello, /u/Theduder89. Thanks for contributing!...
...
4987     Hello, /u/cc_hk. Thanks for contributing! Unfo...
4994     Hello, /u/Realtalkdo3. Thanks for contributing...
4995     Hello, /u/Gransanto102. Thanks for contributin...
4996     Hello, /u/SocialObserver3802. Thanks for contr...
4997     Hello, /u/SthenicFreeze. Thanks for contributi...
Name: text, Length: 182, dtype: object
```

```
poli_dis_2020_df
1      [A reminder for everyone](https://www.reddit.c...
3      [A reminder for everyone](https://www.reddit.c...
19     [A reminder for everyone](https://www.reddit.c...
42     [A reminder for everyone](https://www.reddit.c...
45     [A reminder for everyone](https://www.reddit.c...
...
4909     Your submission was automatically removed.\n\n...
4947     [A reminder for everyone](https://www.reddit.c...
4962     [A reminder for everyone](https://www.reddit.c...
4989     [A reminder for everyone](https://www.reddit.c...
4990     [A reminder for everyone](https://www.reddit.c...
Name: text, Length: 271, dtype: object
```

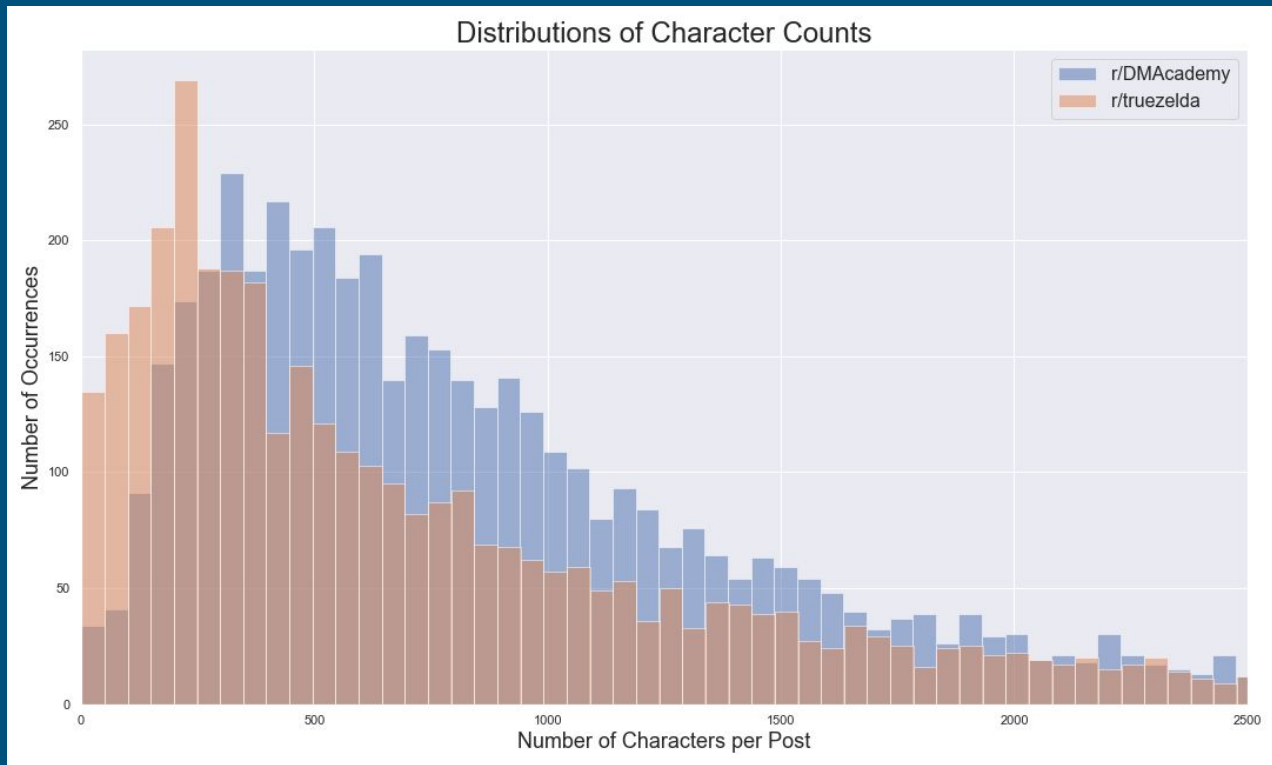
# Stop words



# Text analysis

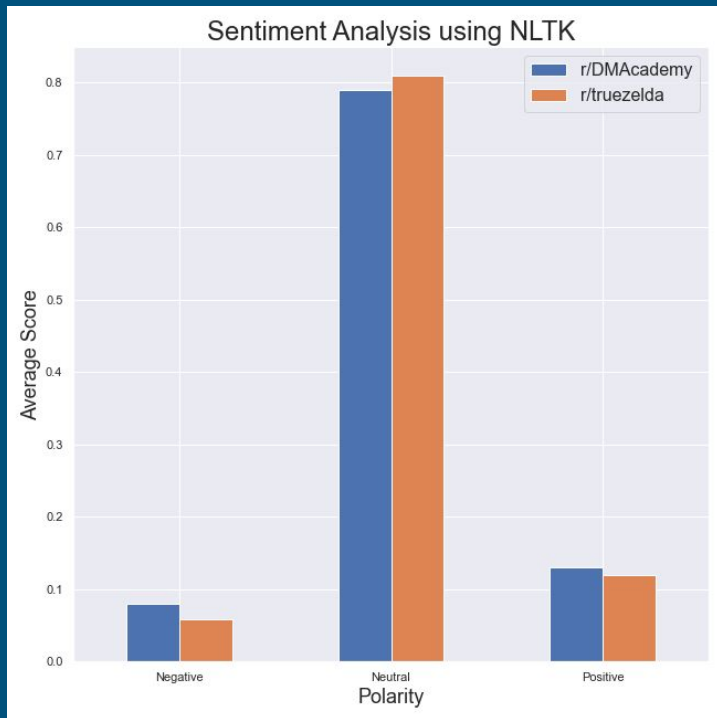


# Text analysis





# Sentiment analysis



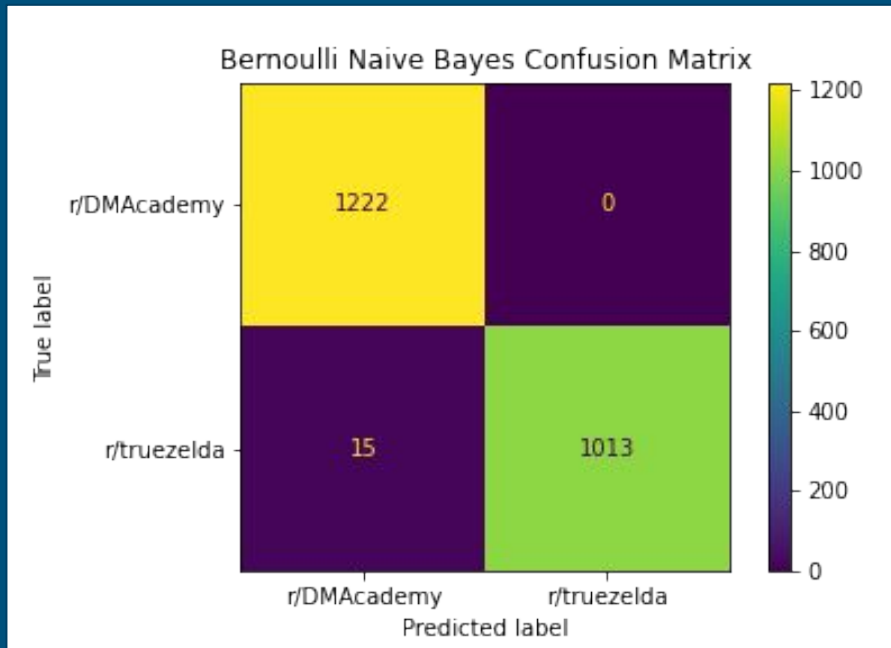
- Used NLTK SentimentIntensityAnalyzer, trained on vader lexicon
- r/DMAcademy was more negative and more positive than r/truezelda
- Both subreddits had more positive posts than negative posts

# Model selection and baseline accuracy

---

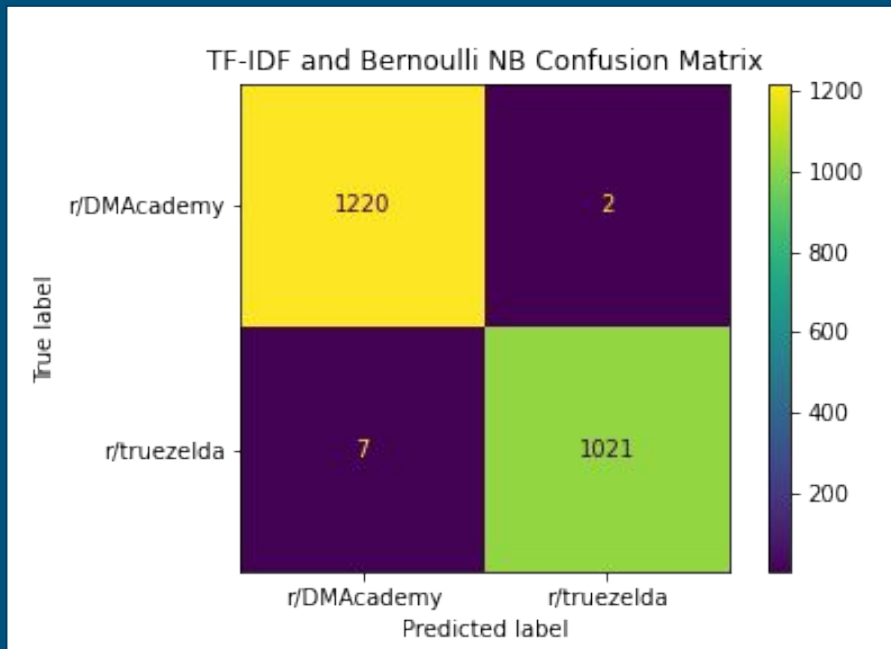
- r/DMAcademy - 54.3%
- r/truezelda - 45.7%
- Bernoulli Naive Bayes - designed for binary classification
  - With Count Vectorizer
  - And with TF-IDF Vectorizer
- Support Vector Machines
  - With TF-IDF Vectorizer

# CVec and Bernoulli Naive Bayes predictions



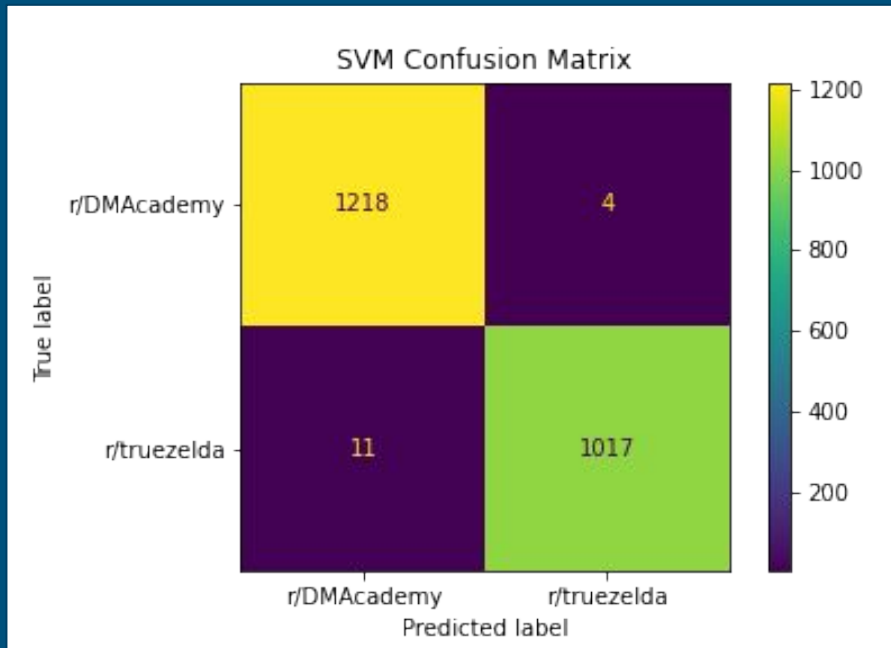
- Started with Count Vectorizer
- Minimal hyperparameter tuning required
- Optimal alpha value of 0.0069519279617756054

# TF-IDF and Bernoulli Naive Bayes predictions



- TF-IDF Vectorizer reduced errors
- Minimal hyperparameter tuning required
- Optimal alpha value of 0.0069519279617756054

# SVM predictions



- Used TF-IDF Vectorizer
- Minimal hyperparameter tuning required
- Optimal alpha value of 6.2105263157894735

# Model evaluation

- Each of these models performed much better than our baseline!
- Accurately predicted which subreddit a post came from 98-99% of the time
- High precision and recall means models predicted true positives accurately (positive was r/truezelda)
- High f1 due to good balance between precision and recall scores

CVec and Bernoulli NB Classification Report

	precision	recall	f1-score
r/DMAcademy	0.987874	1.000000	0.993900
r/truezelda	1.000000	0.985409	0.992651
accuracy	0.993333	0.993333	0.993333

TF-IDF and Bernoulli NB Classification Report

	precision	recall	f1-score
r/DMAcademy	0.994295	0.998363	0.996325
r/truezelda	0.998045	0.993191	0.995612
accuracy	0.996000	0.996000	0.996000

CVec and SVM Classification Report

	precision	recall	f1-score
r/DMAcademy	0.991050	0.996727	0.993880
r/truezelda	0.996082	0.989300	0.992679
accuracy	0.993333	0.993333	0.993333

# Possible next steps

---

- Print out misclassified posts to investigate potential sources of error
- Explore documentation to understand why SVM makes more mistakes than Bernoulli Naive Bayes
- Apply the same workflow to predict what year a post came from
  - Under development, look forward to seeing it in my Jupyter notebooks!

# Notes to self

---

Precision =  $tp / (tp + fp)$

Recall =  $tp / (tp + fn)$

F1 =  $precision * recall / (precision + recall)$