

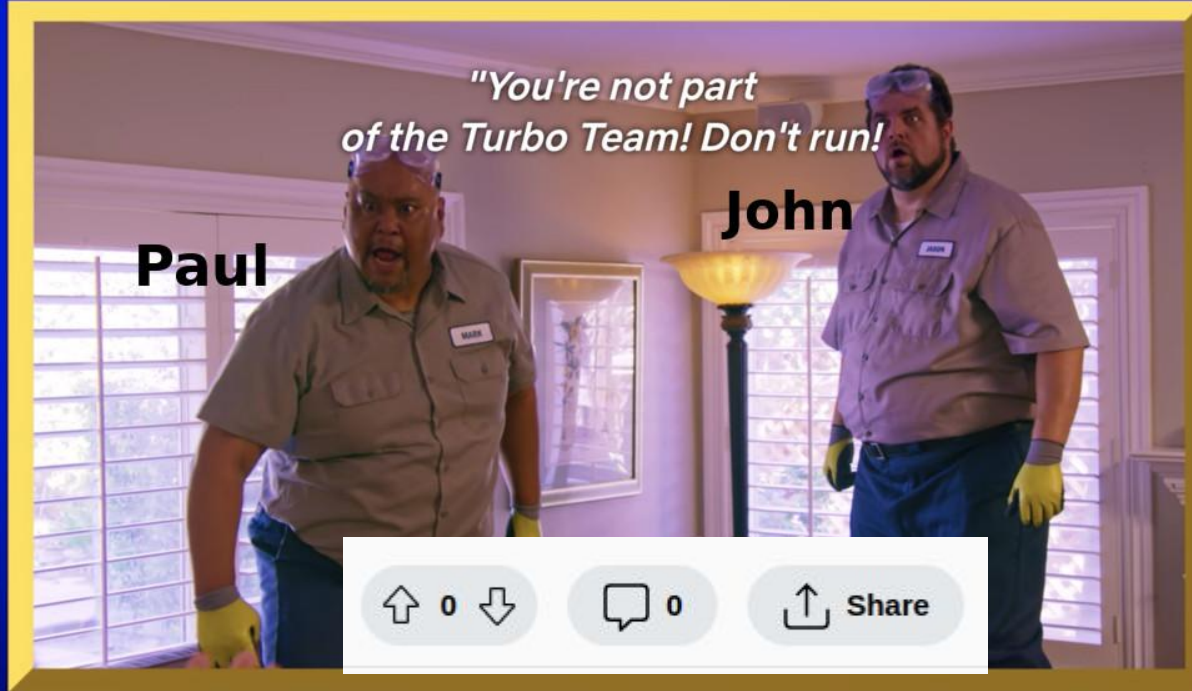
Has this ever happened to you?



George MITCH BRYANT
CALL ME NOW!

1-888-555-MITCH

Has this ever happened to you?



George MITCH BRYANT
CALL ME NOW!

1-888-555-MITCH

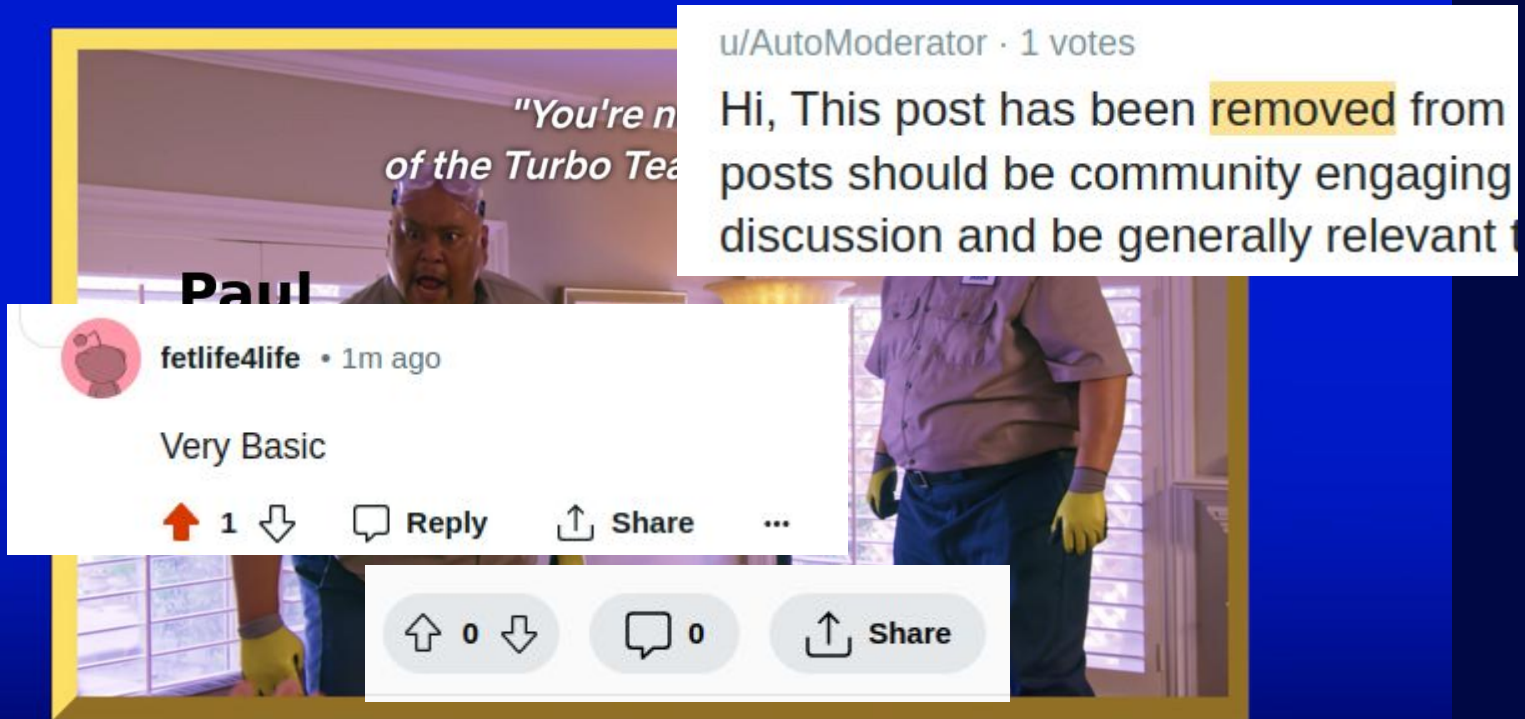
Has this ever happened to you?



George MITCH BRYANT
CALL ME NOW!

1-888-555-MITCH

Has this ever happened to you?



George MITCH BRYANT
CALL ME NOW!

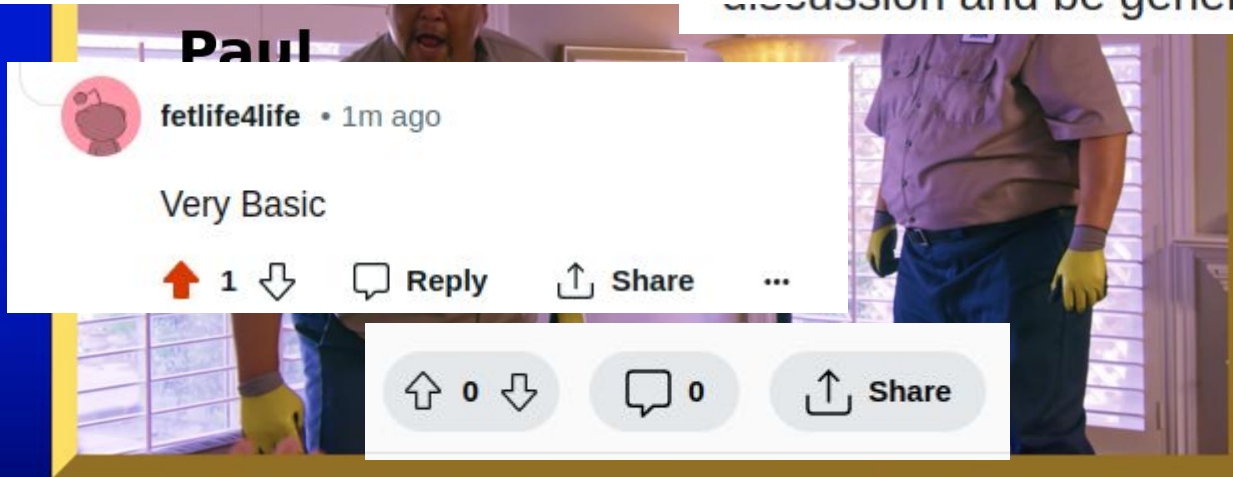
1-888-555-MITCH

Has this ever happened to you?

You try to find a reddit that will accept you
and your memes, but instead...
rejection, bullying, and 0 upvotes

toModerator · 1 votes

This post has been removed from
its should be community engaging
discussion and be generally relevant t



George MITCH BRYANT
CALL ME NOW!

1-888-555-MITCH

Has this ever happened to you?

You try to find a reddit that will accept you and your memes, but instead... rejection, bullying, and 0 upvotes

toModerator · 1 votes

This post has been removed from its should be community engaging discussion and be generally relevant t

Paul



fetlife4life · 1m ago

Very Basic



1



Reply



Share



George

NLP to identify key cultural features through text to camouflage writing style to win approval of reddit groups. Using Beatles Subreddits as a case study.

Is there a notable difference among Beatles Subreddits?

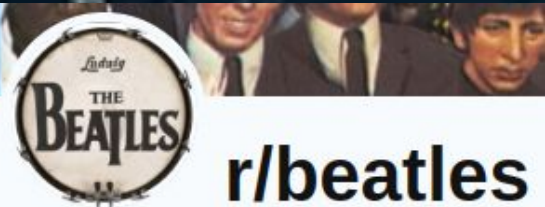
The Subs:

r/Beatles
r/TheBeatles
r/BeatlesCircleJerk

The Questions:

1. How accurately can different classification models correctly identify which subreddit text came from?
2. Do different subs prefer different Beatles?
3. Which sub has the most bullies/NSFW?

Is there a notable difference among Beatles Subreddits?



r/beatles

The Beatles

Welcome to The Beatles!

222K

Members

103

Online

Top 1%

Rank by size ↗



r/TheBeatles

The Beatles

The Beatles were an English rock group from 1960-1970. Their influence on the world and today's music is incredible. This subreddit is for discussion, news, memorabilia, and/ or anything else relating to the group. :)

66K

Members

18

Online

Top 2%

Rank by size ↗



r/beatlesciclejerk

Beat The Meatles

Revolution 9 sucks amirite??? Punchy, Granny, Who?, The Drummer. r/beatles has no sense of humor r/TheBeatles has a 65+ age requirement r/pinkfloydcirclejerk banned me r/radioheadcirclejerk who?

98K

Members

29

Online

Data:

500 posts from each sub
+ 10 comments from each post

Is there a notable difference among Beatles Subreddits?

The Subs:

r/Beatles

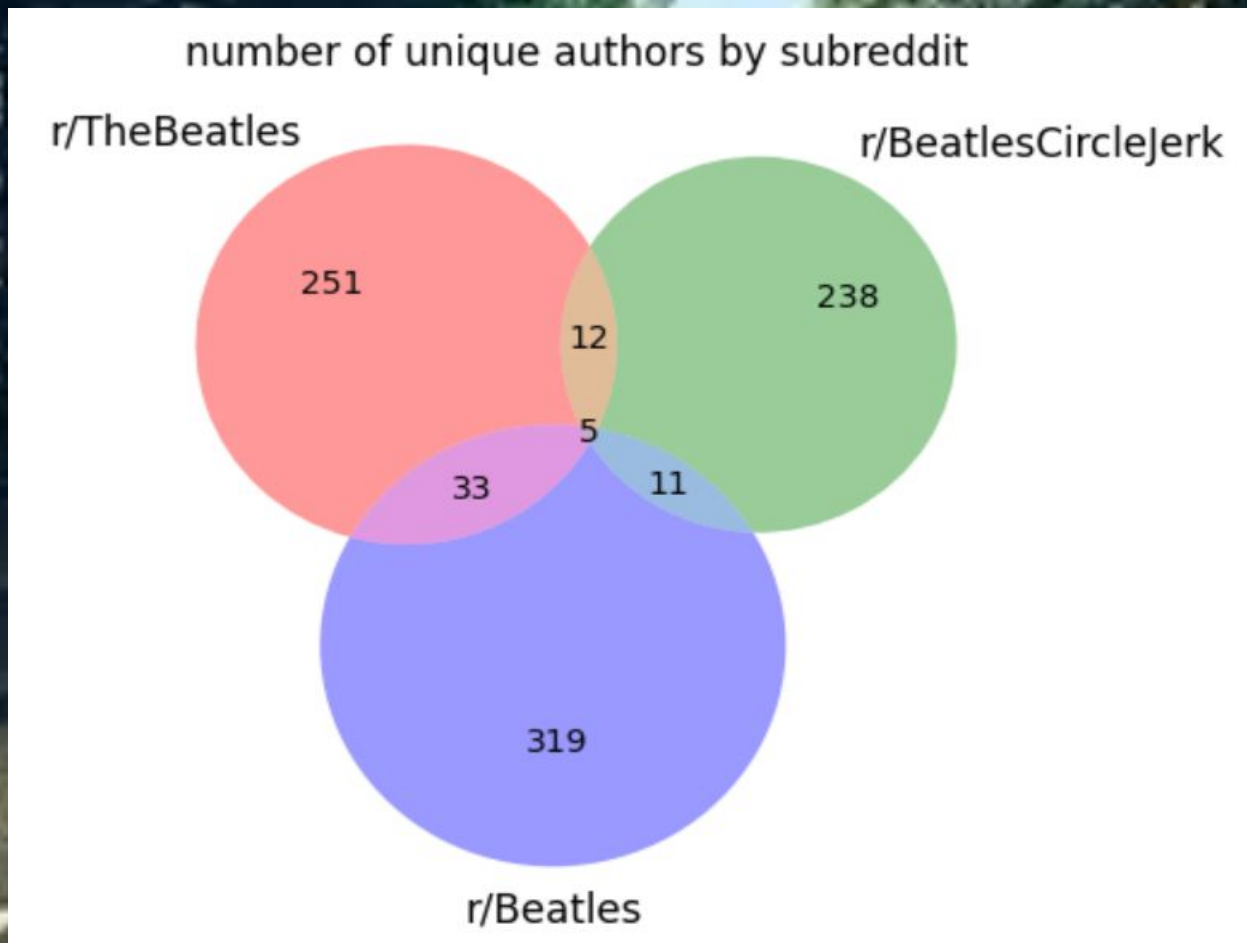
r/TheBeatles

r/BeatlesCircleJerk

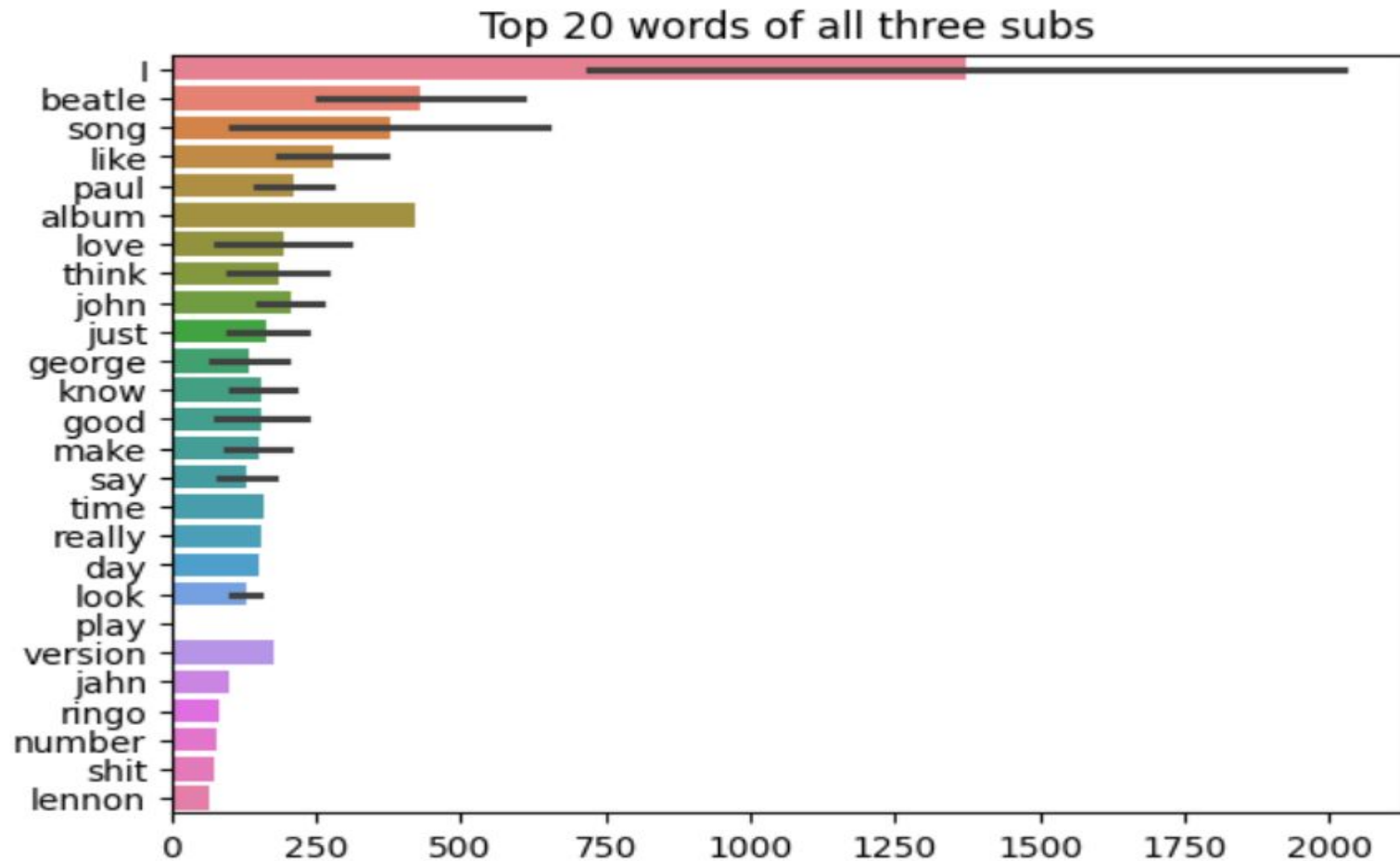
The Method:

1. Capture posts with comments to gain essence of each subreddits culture
2. Look for unique and popular words
3. Fit various classification models to identify contents origin

Analysis: common authors across groups



Top words using CountVectorizer and Lemmatization

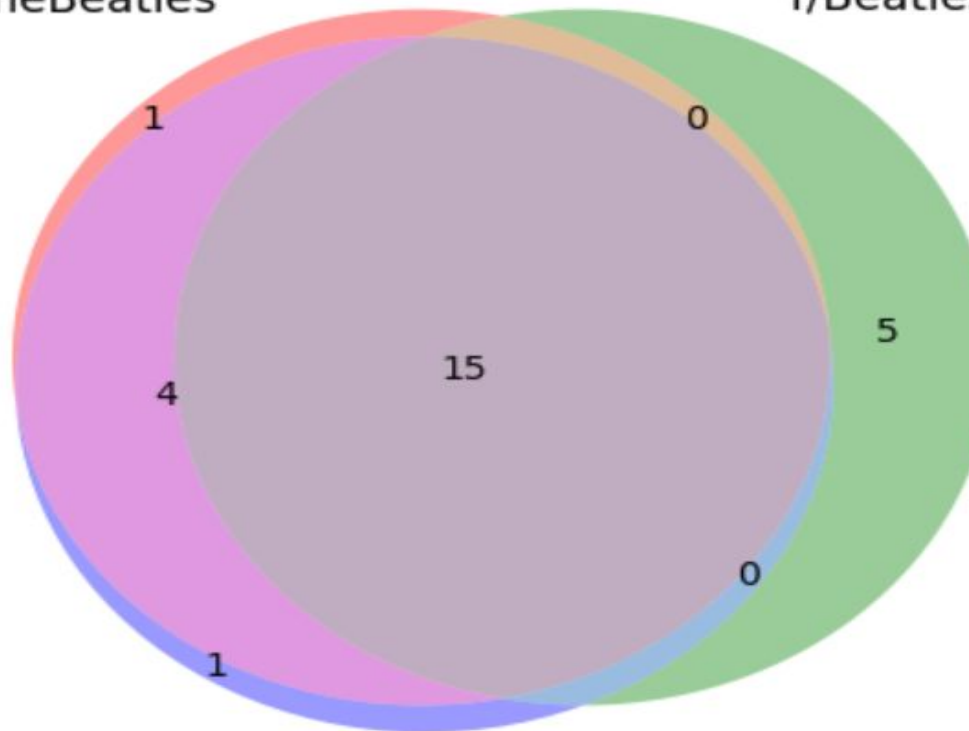


Top words using CountVectorizer and Lemmatization

number of unique top 20 words by subreddit

r/TheBeatles

r/BeatlesCircleJerk



'jahn',
'lennon',
'number',
'ringo',
'shit'

r/Beatles

Who are the most popular beatles across each sub.

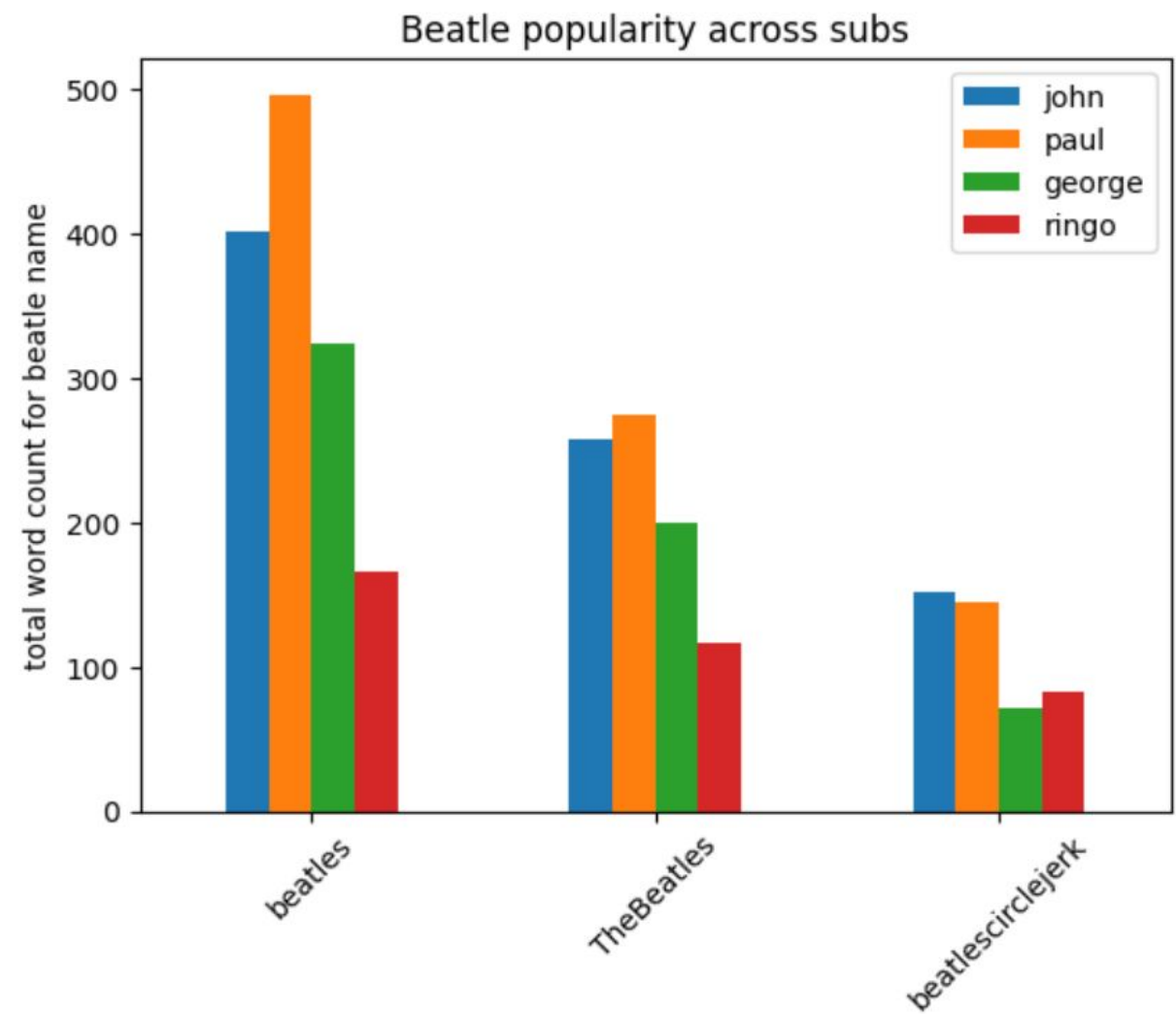
The Subs:

r/Beatles and
r/TheBeatles

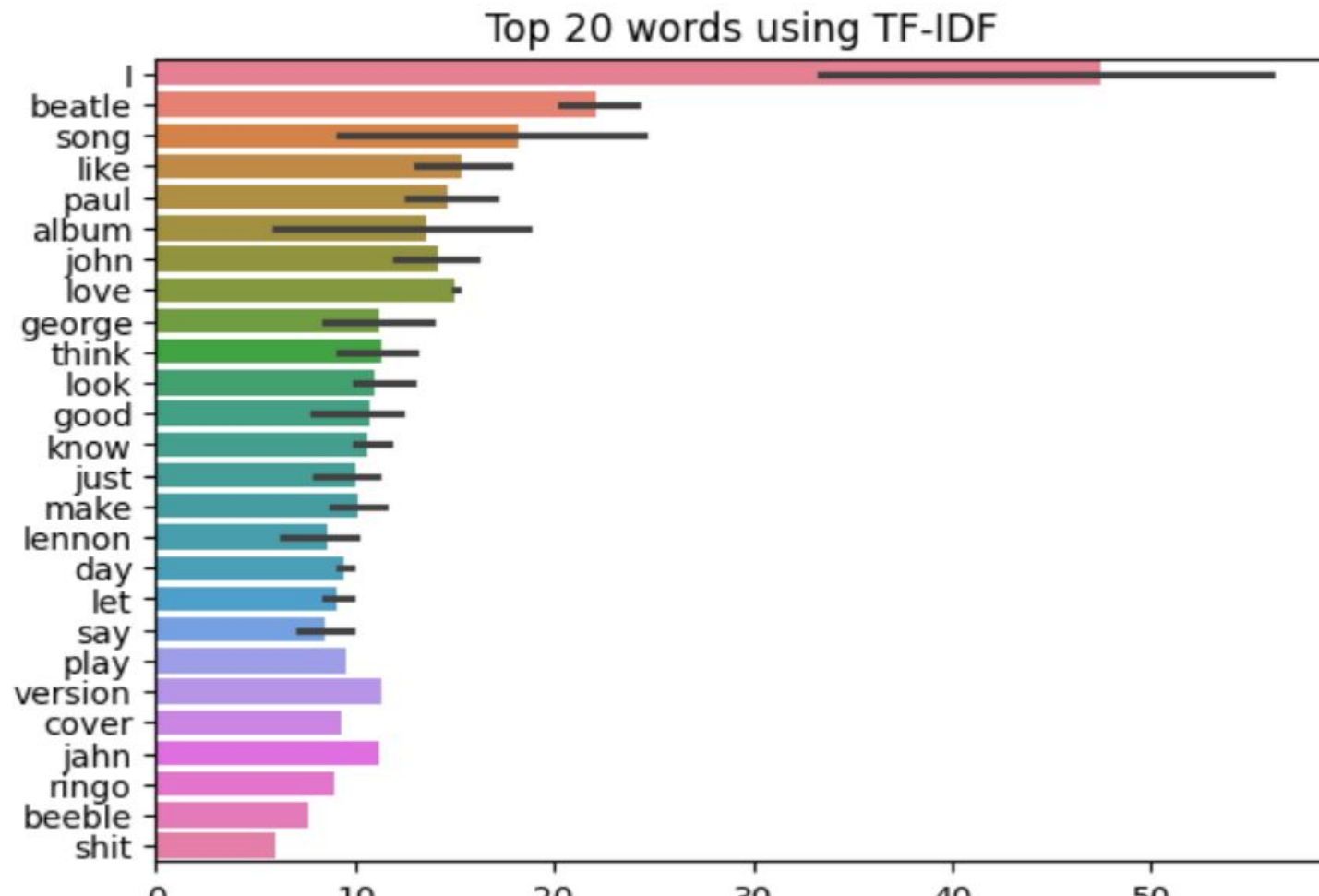
Paul is more popular than John and Ringo is not even in the top 20 words.

r/BeatlesCircleJerk:

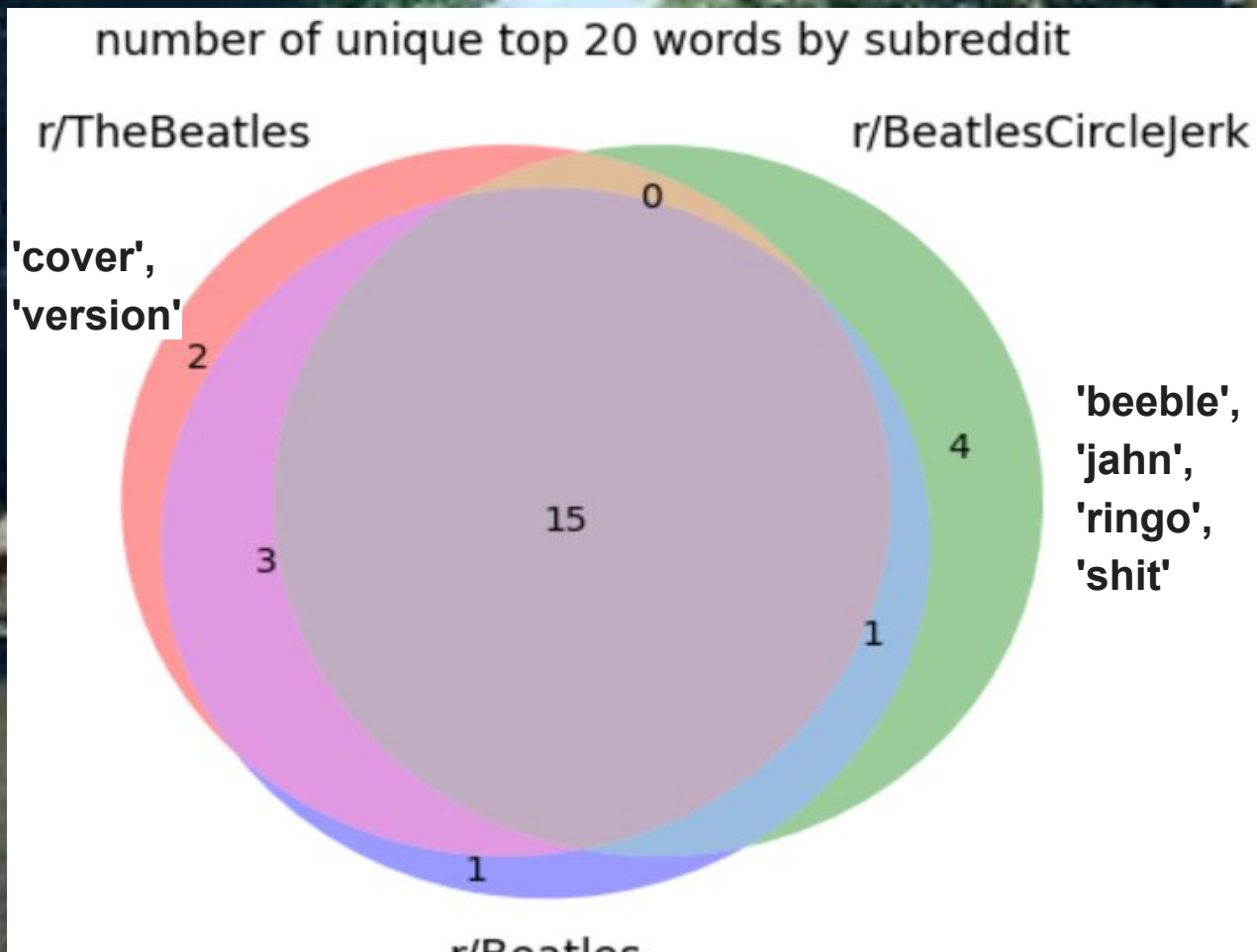
John is more popular than Paul, and Ringo is more popular than George



Top words using TF-IDF and Lemmatization

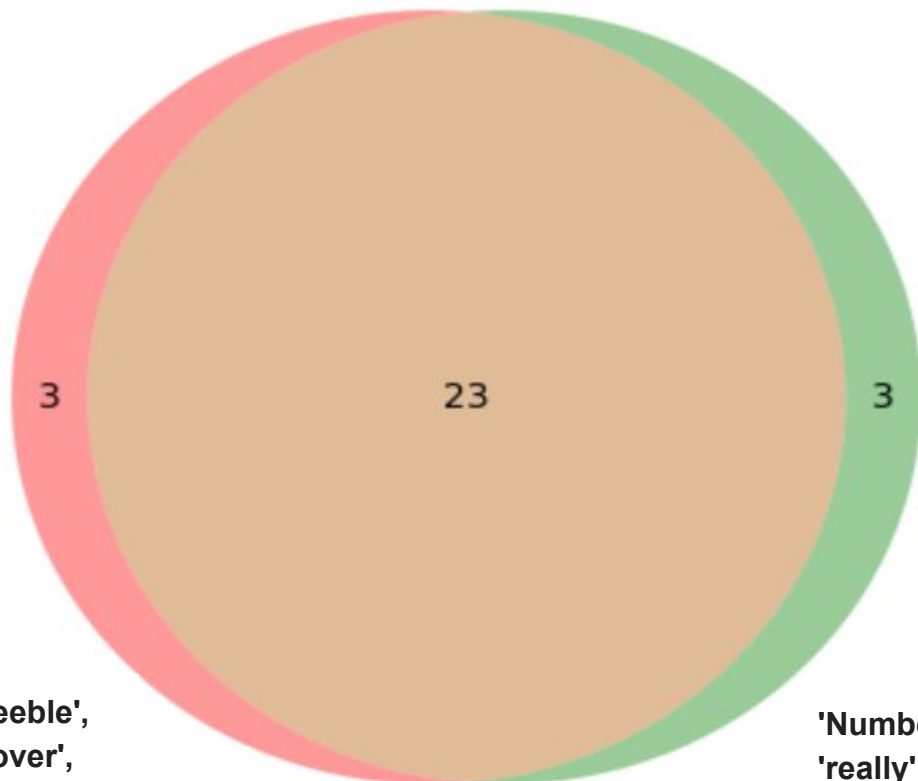


Top words using TF-IDF and Lemmatization



Top words further analysis

number of unique top words cvect vs tfidf



'beebie',
'cover',
'let'

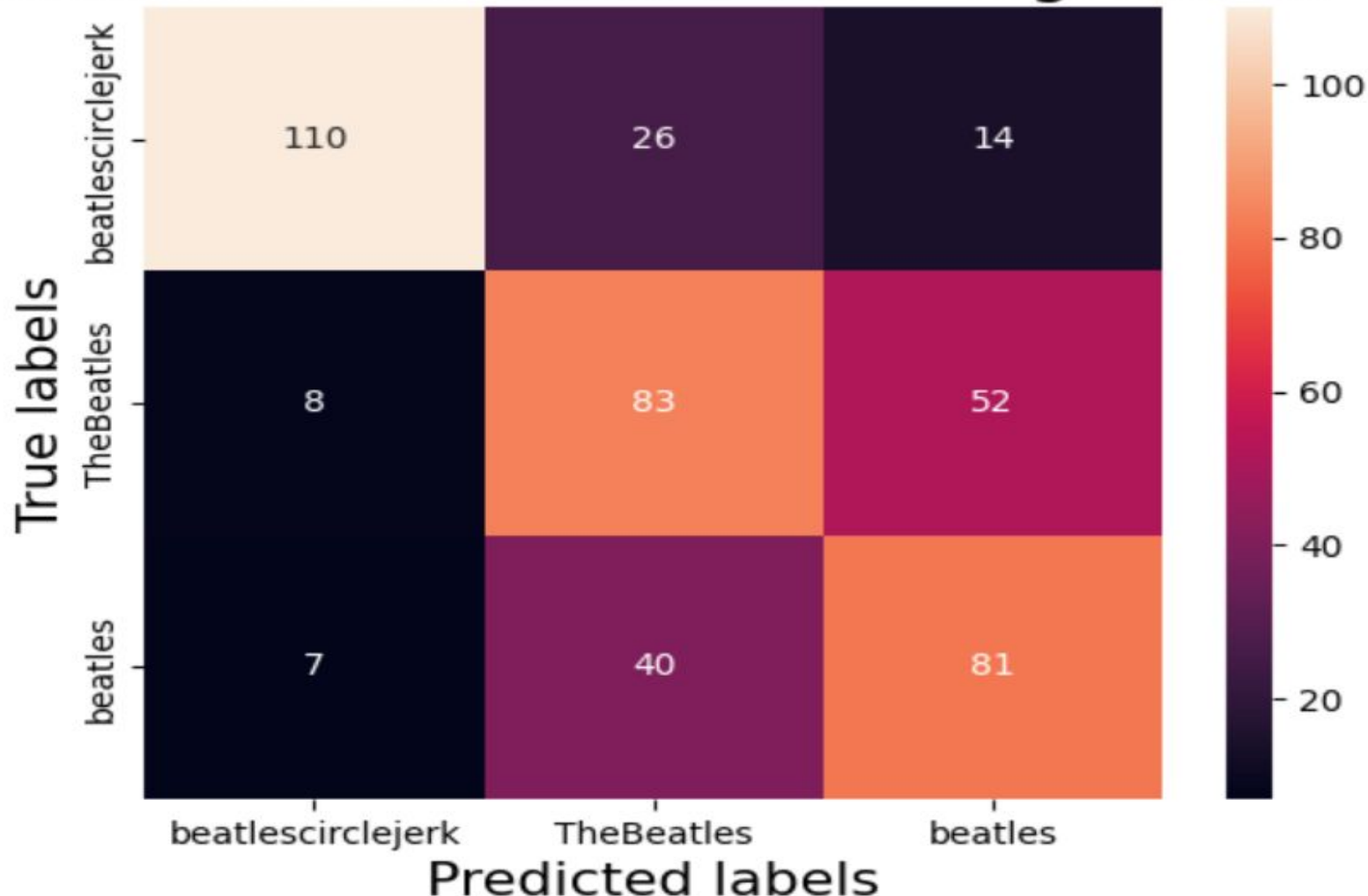
'Number',
'really',
'time'

TF IDF CVec

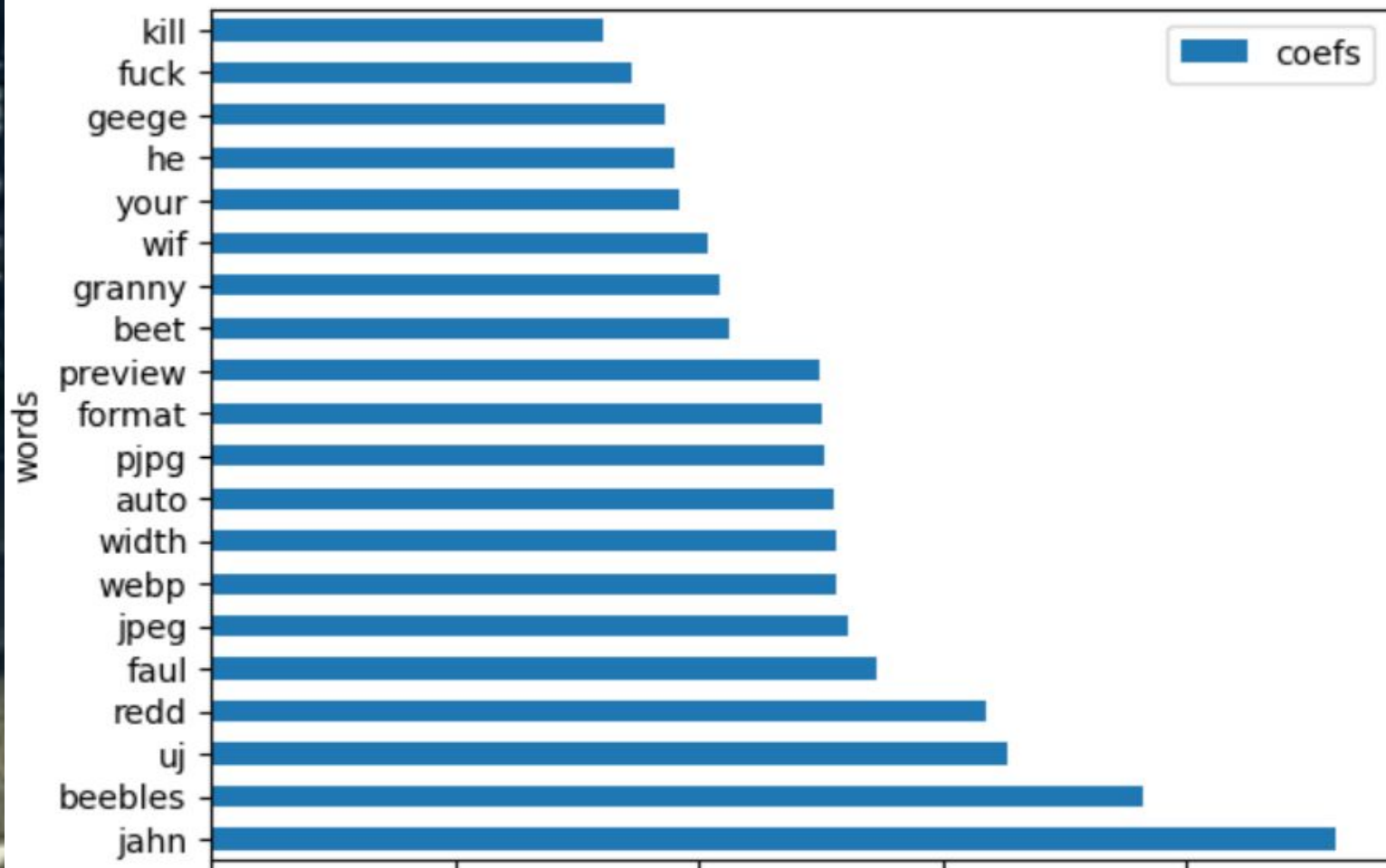
Top Models with minimal optimization:

	LogRegression	randomforest	knn_class	svc
model	LogRegression	randomforest	knn_class	svc
train_score	0.926784	0.999073	0.653383	0.982391
test_score	0.657957	0.63658	0.593824	0.655582
f1_score	0.655221	0.635347	0.595638	0.650379

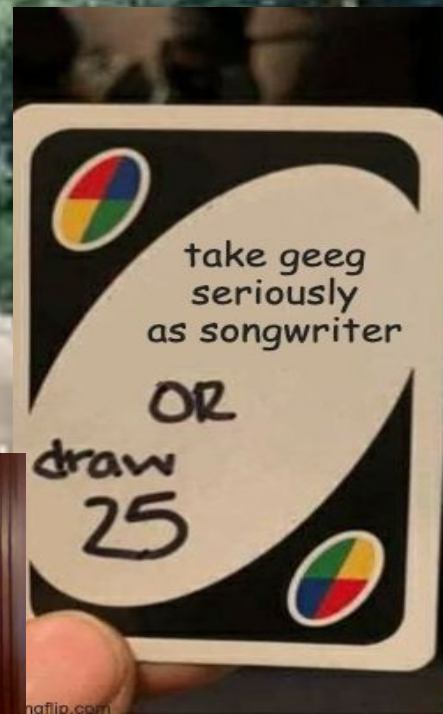
Confusion Matrix of tf-idf + logression



Analysis through tf-idf vectorization and logistic regression



Hire professional Shark Tank writer to create data driven memes



GRANNY: BEST I CAN DO WILD HONEY PIE

Conclusion and further investigation:

Further Study:

1. Further analysis: Use data to determine best time to post to subs
2. Test popularity of data driven memes on Beatles subs
3. Create model to better differentiate between r/Beatles and r/TheBeatles
4. Improve model performance overall through further hyper parameterization
5. Review findings, be accepted as fellow redditor with data science