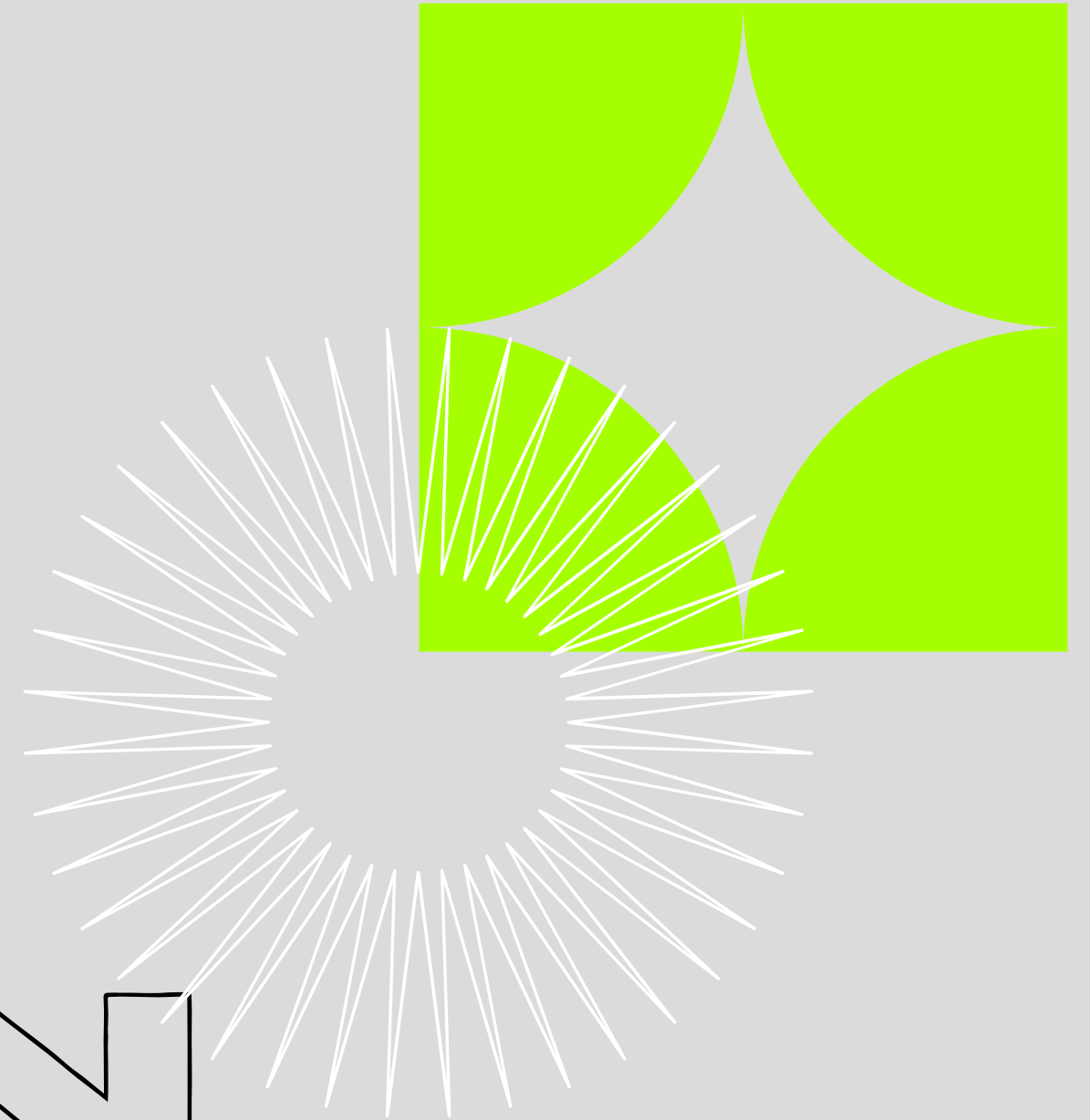M. DOWLING | DSI IMMERSIVE

RELIGION
CULT

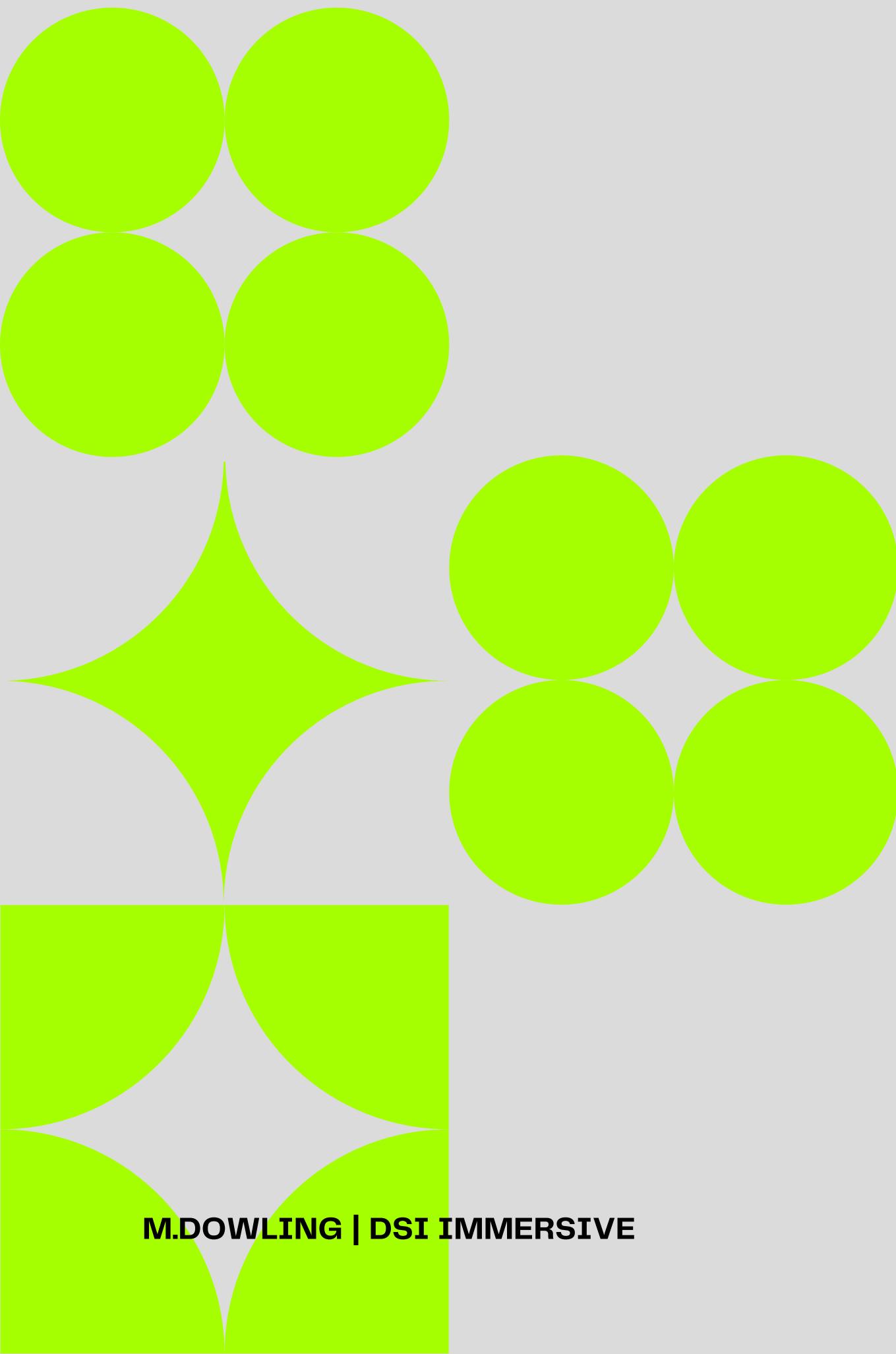## A NLP CLASSIFICATIONN ANALYSIS

BACKGROUND

1931

- term cult originated in the sociology study of religion

1970

- Rise of the secular anti-cult movement
- term 'cult' replaced with 'religious movement'

Today

- need of definitions that differentiate destructive/ dangerous groups from those that are benign
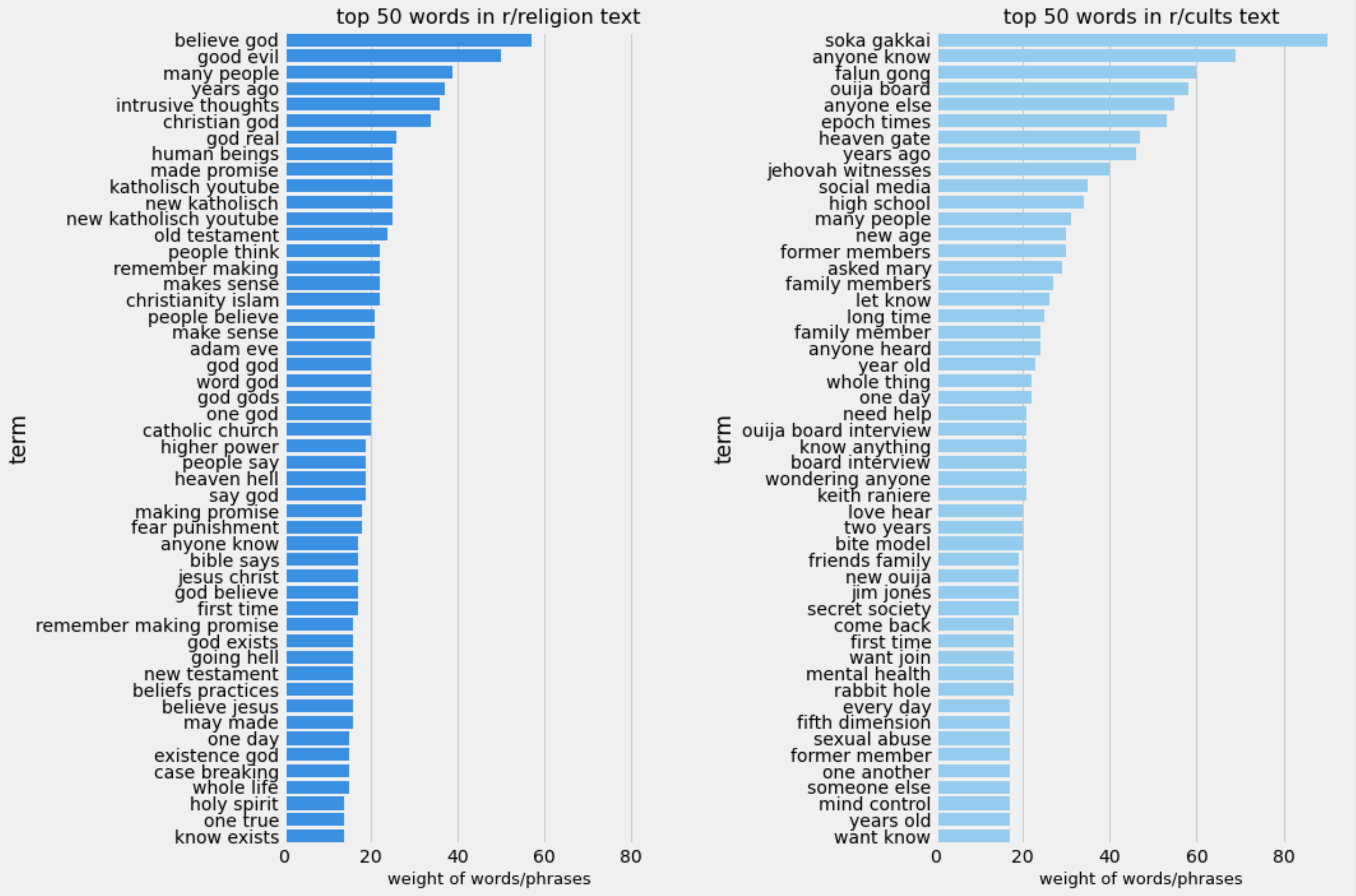
PROBLEM STATEMENT

As better definitions for a dangerous cult and a new religious movement struggle to come to an official understanding, Our client, who prefers to be kept r/anonymous, has requested to know if a dangerous group can be differentiated from a more benign group through social media posts so they can identify the rise of a dangerous group prior to the health and wellbeing of individuals is at risk. In order to accomplish this, we will use Reddit to collect posts under the two conflicting umbrella terms r/religion and r/cults and identify commonalities and differences within the vocabulary and sentiment of each post. Through Natural Language processing and classification modeling our goals are to determine how distinct these groups are and what type of language style and vocabulary helps to predict a post subreddit.

SENTIMENT

Sentiment by Subreddit
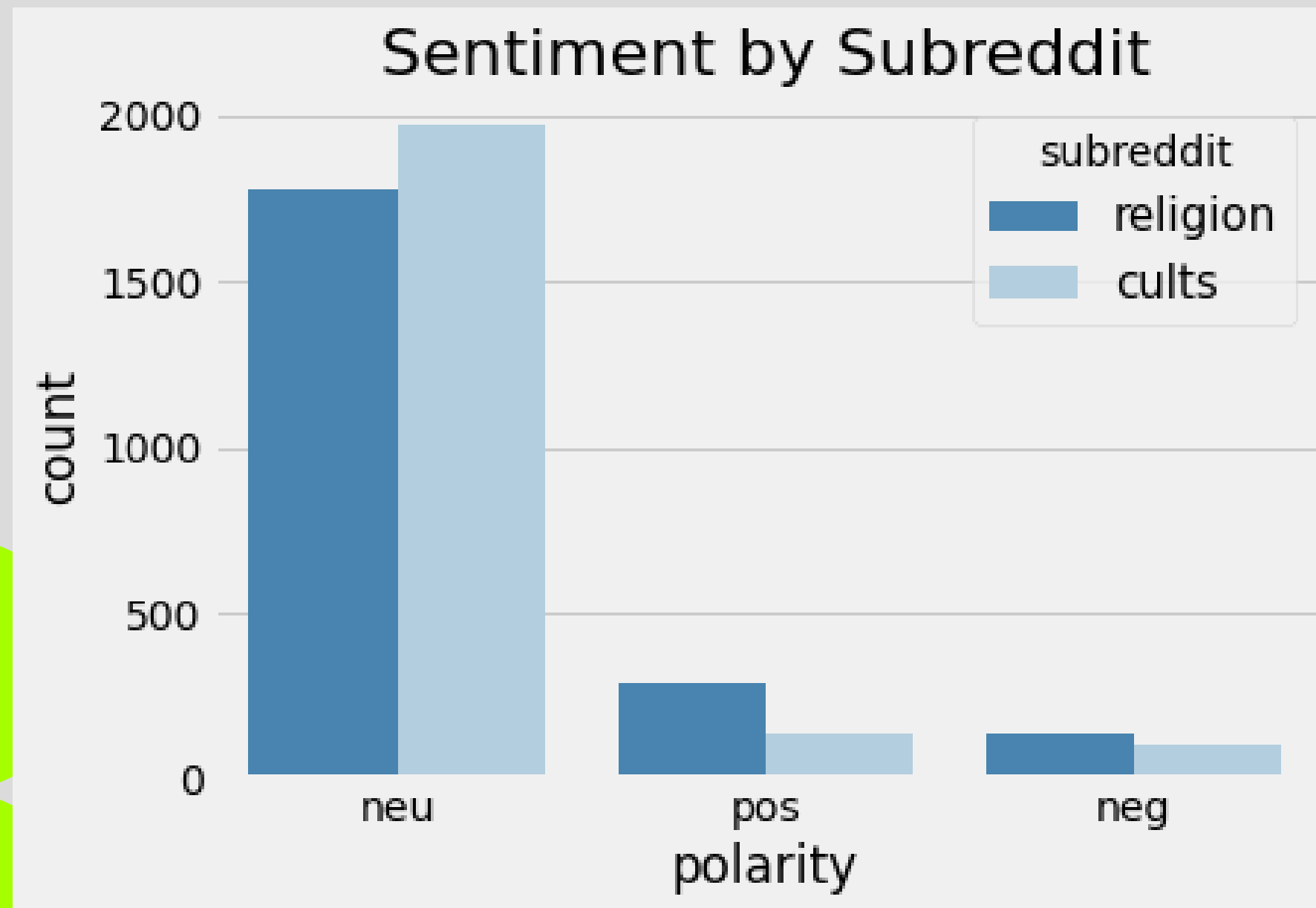
- Mostly neutrally defined statements

- religion has more negative and positive statements.

BIGRAMS

top 50 words in r/religion text

term	
believe god	
good evil	
many people	
years ago	
intrusive thoughts	
christian god	
god real	
human beings	
made promise	
katholisch youtube	
new katholisch	
new katholisch youtube	
old testament	
people think	
remember making	
makes sense	
christianity islam	
people believe	
make sense	
adam eve	
god god	
word god	
god gods	
one god	
catholic church	
higher power	
people say	
heaven hell	
say god	
making promise	
fear punishment	
anyone know	
bible says	
jesus christ	
god believe	
first time	
remember making promise	
god exists	
going hell	
new testament	
beliefs practices	
believe jesus	
may made	
one day	
existence god	
case breaking	
whole life	
holy spirit	
one true	
know exists	

weight of words/phrases

top 50 words in r/cults text

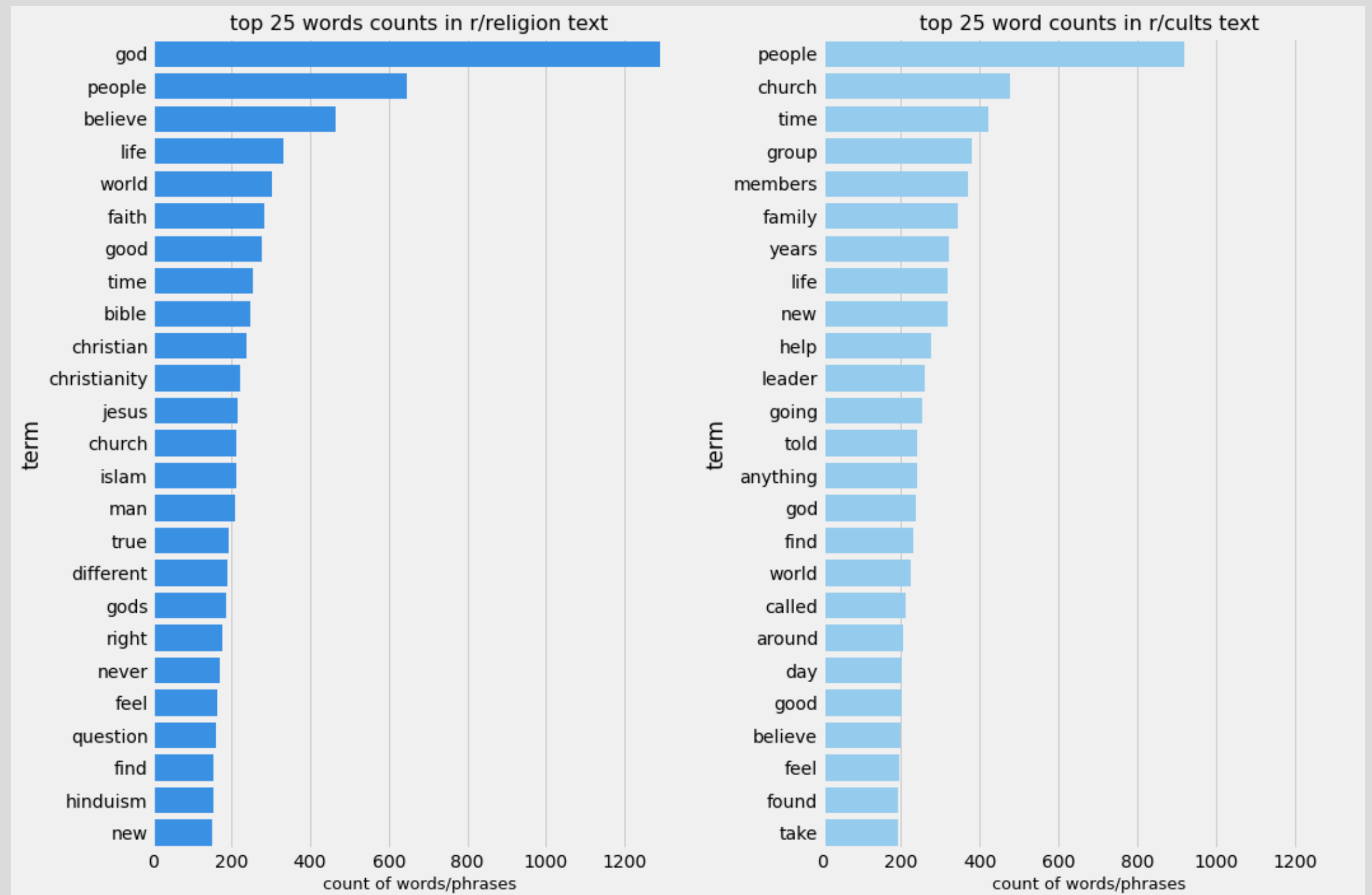| term	
soka gakkai	
anyone know	
falun gong	
ouija board	
anyone else	
epoch times	
heaven gate	
years ago	
jehovah witnesses	
social media	
high school	
many people	
new age	
former members	
asked mary	
family members	
let know	
long time	
family member	
anyone heard	
year old	
whole thing	
one day	
need help	
ouija board interview	
know anything	
board interview	
wondering anyone	
keith raniere	
love hear	
two years	
bite model	
friends family	
new ouija	
jim jones	
secret society	
come back	
first time	
want join	
mental health	
rabbit hole	
every day	
fifth dimension	
sexual abuse	
former member	
one another	
someone else	
mind control	
years old	
want know	

weight of words/phrases

M.DOWLING | DSI IMMERSIVE

EXPLORATORY ANALYSIS

The bar graphs to the right are displaying the top 25 most common words in each respective subreddit. Two things to note:

- There are many similarities in terms
- There is an 'otherness' aspect in the r/cults graph using more colliquical terms that any social group might use.


top 25 words counts in r/religion text

top 25 word counts in r/cults text

EXPLORATORY ANALYSIS

The Venn diagram displays the amount of overlapping vocabulary within the top 25 most frequent terms. The 11 terms that are overlapping could be an indication that differentiating these posts in modeling will be difficult.

top 25 common words in text

14   11   14

r/religion   r/cults
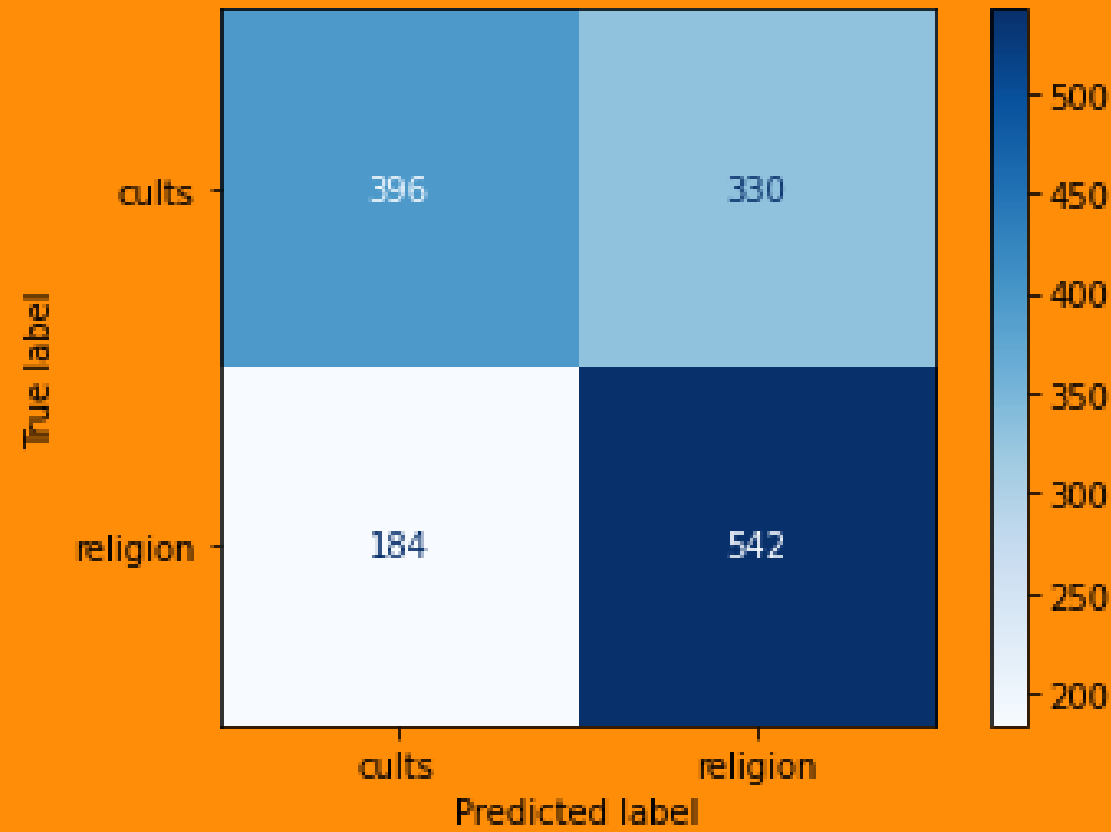
MODELING

KNN & CVEC
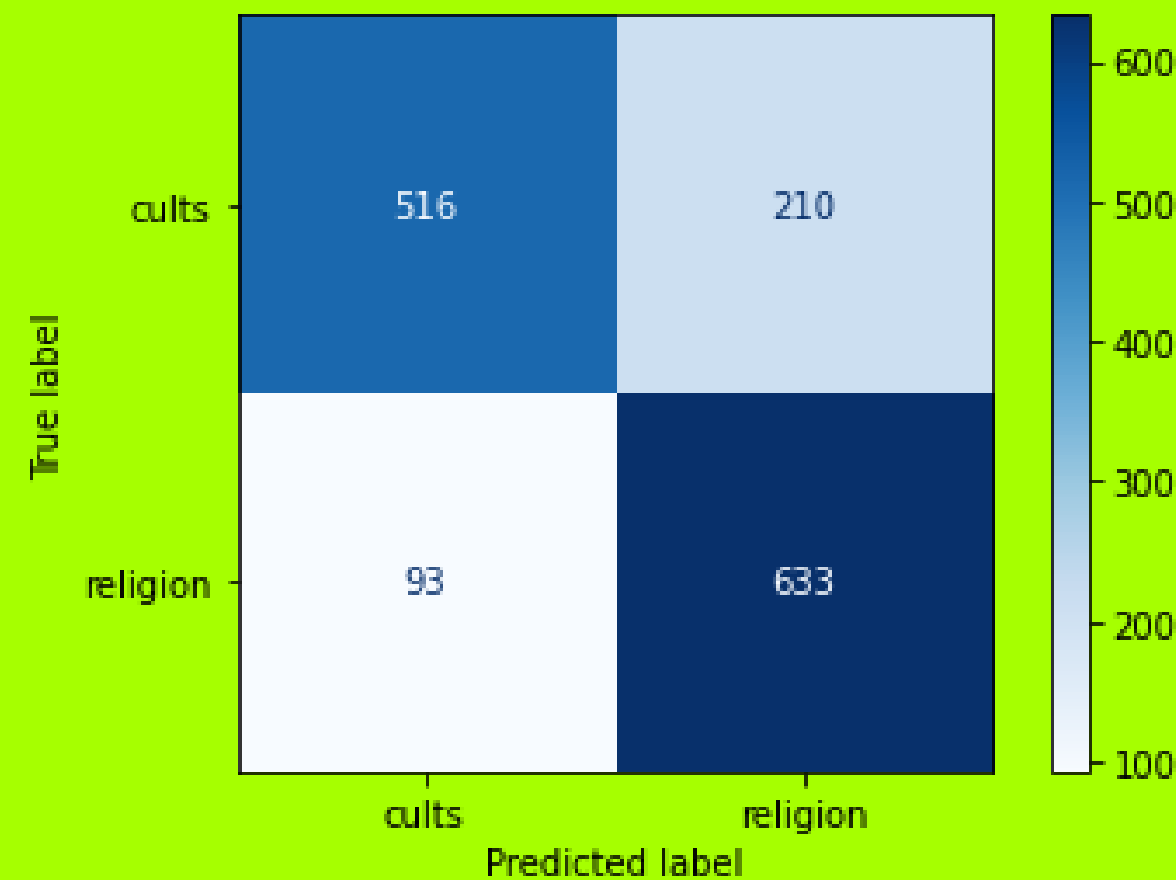

Random Forest & CVEC
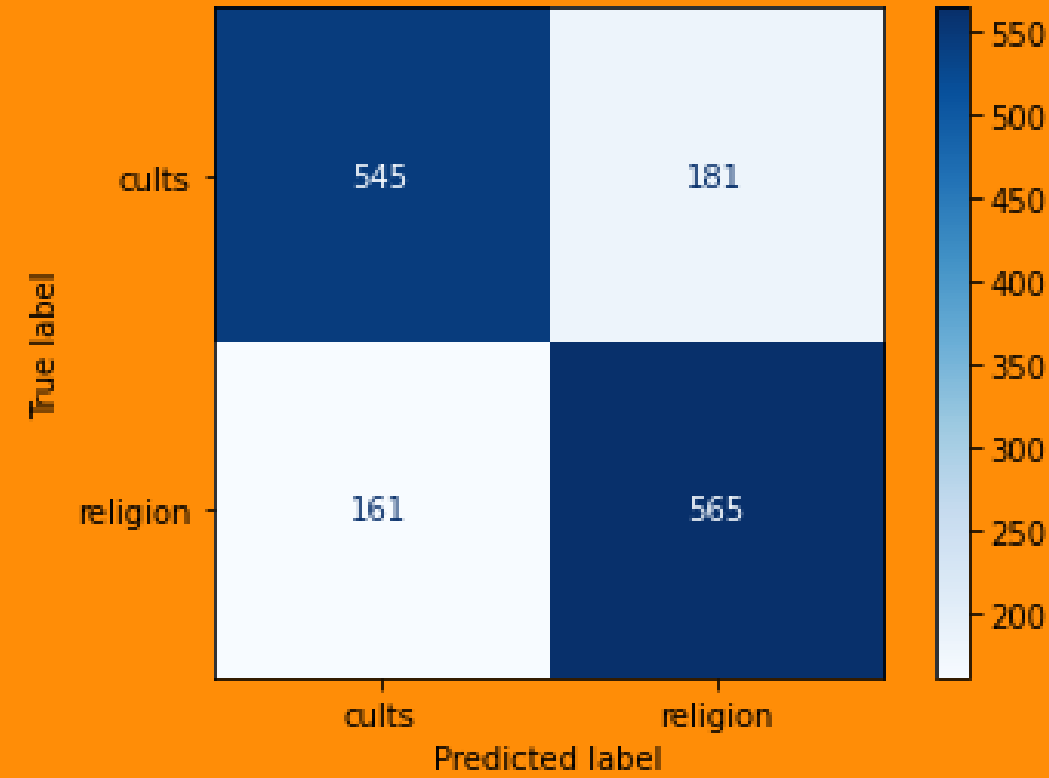
Multinominal Naïve Bayes
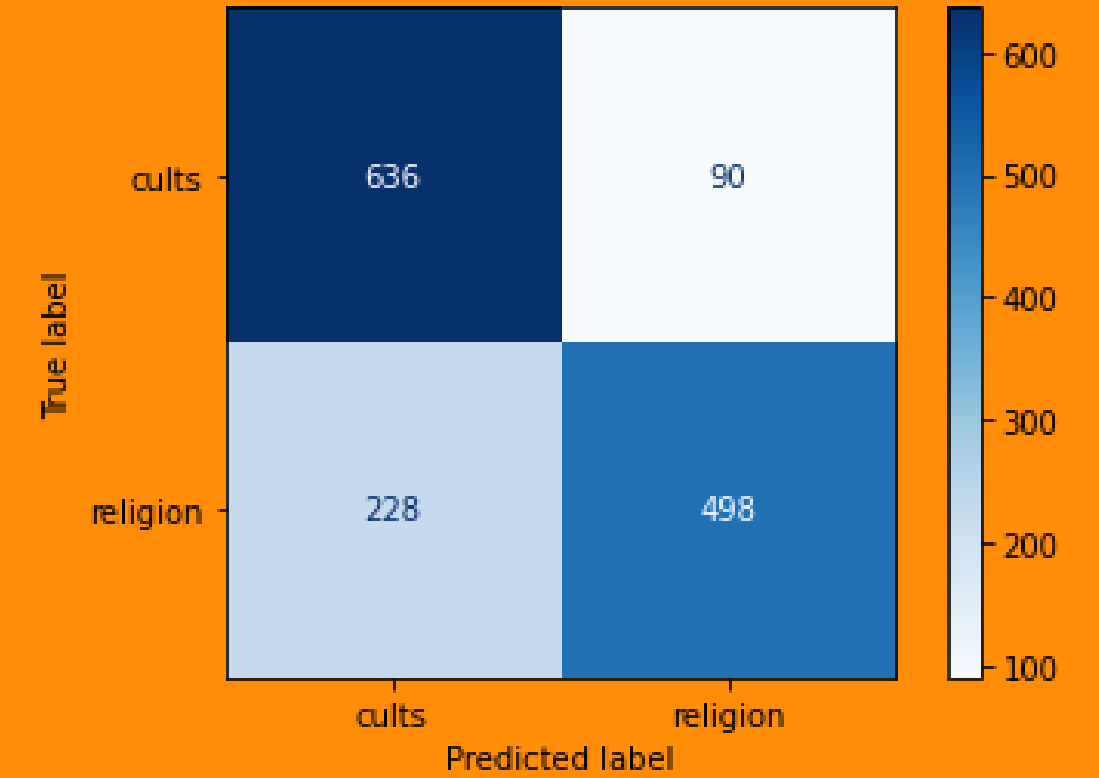
Model	Transformer	Train Accuracy	Test Accuracy	Recall	Percision
KNN	cvec	0.76	0.64	0.55	0.68
Random Forest	cvec	0.96	0.76	0.75	0.77
Naïve Bayes	cvec	0.81	0.78	0.88	0.74

MODELING

KNN & TVEC
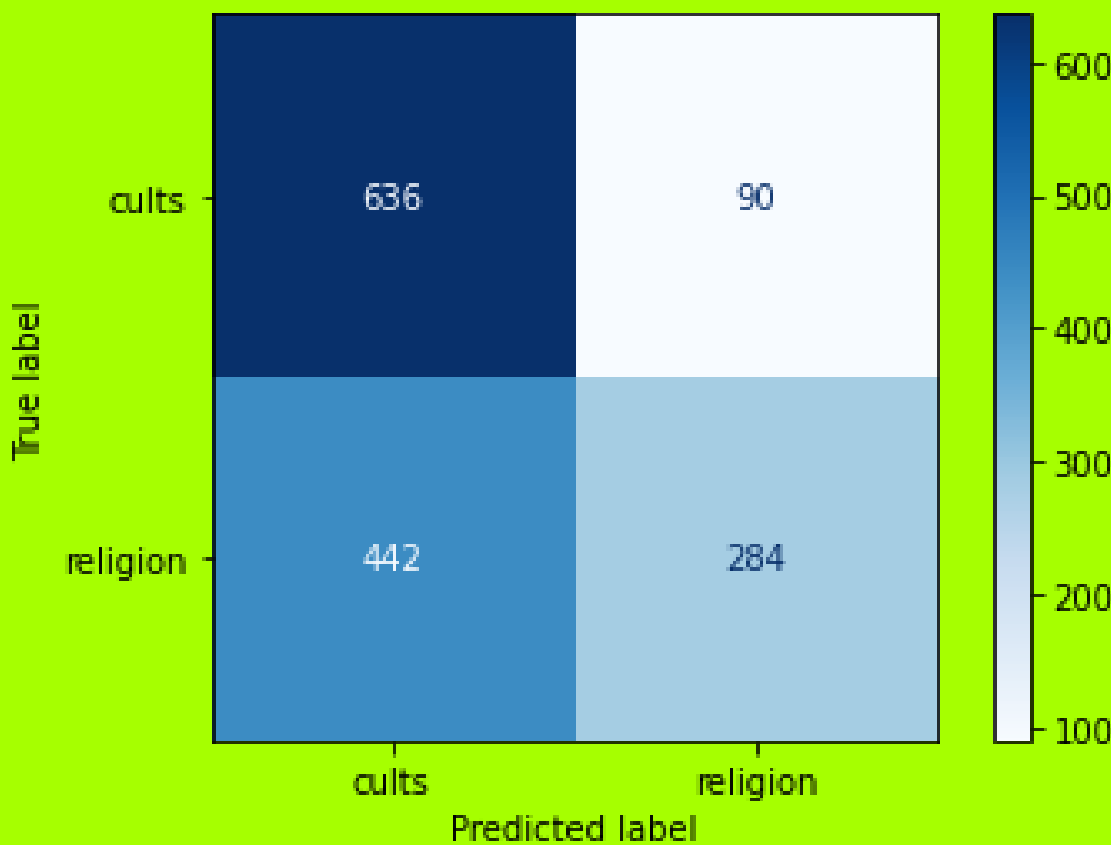
Random Forest & TVEC

Naïve Bayes MN & TVEC
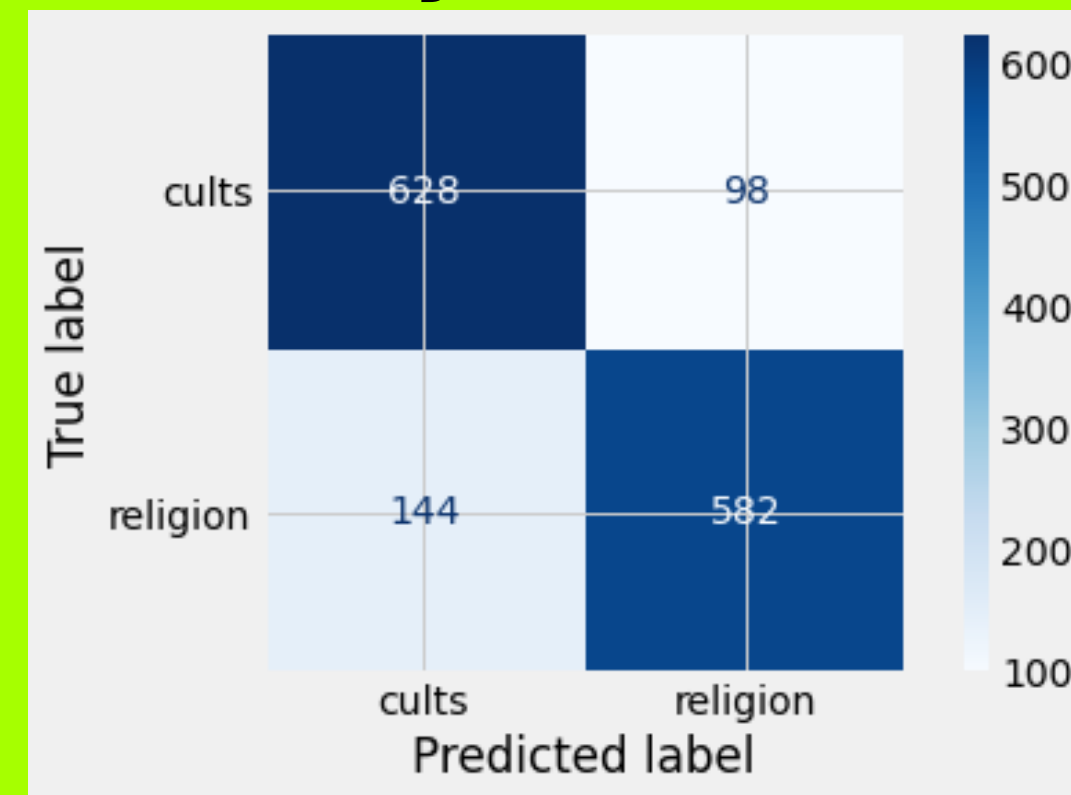
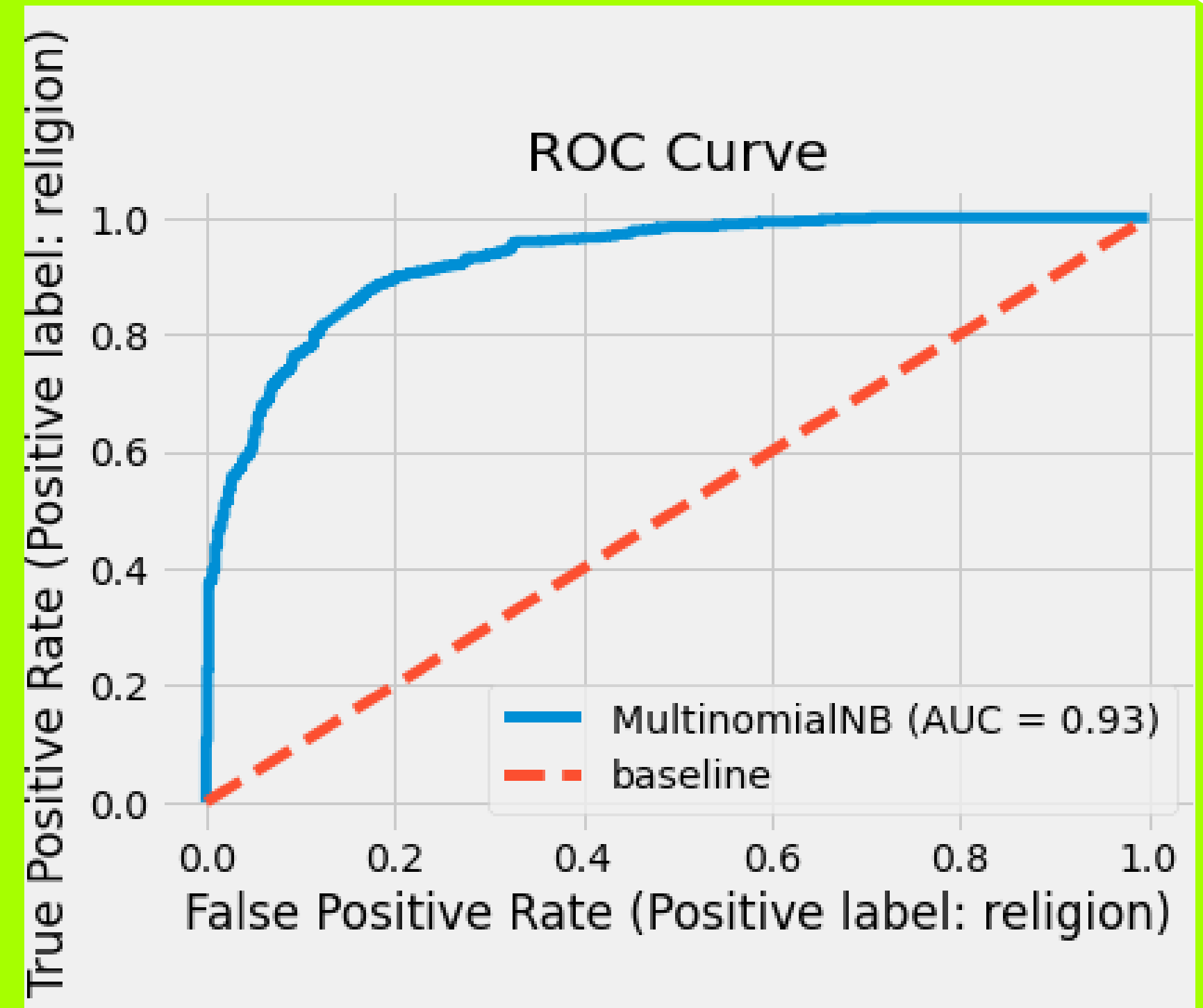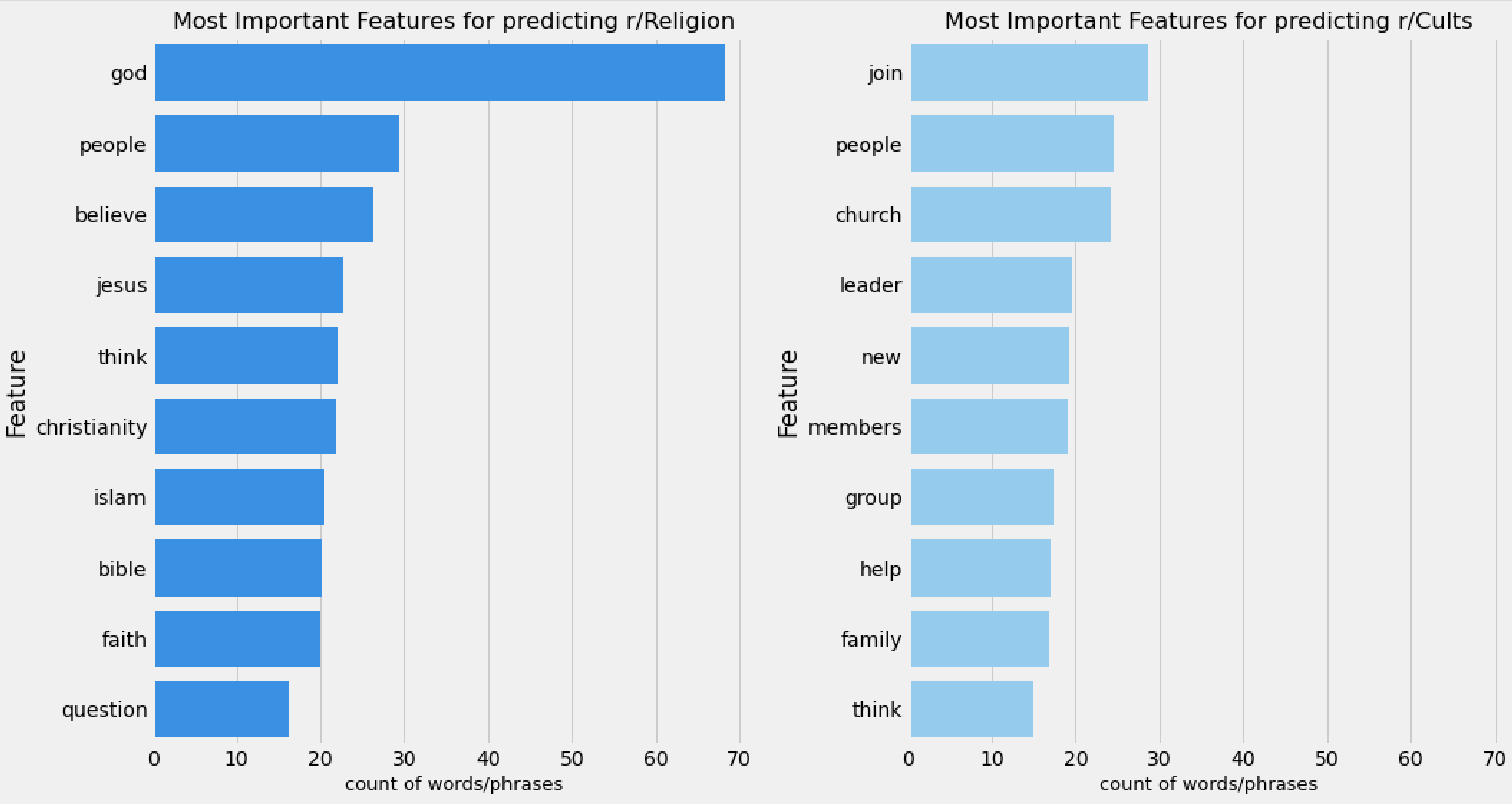| Model	Transformer	Train Accuracy	Test Accuracy	Recall	Percision
KNN	tvec	0.95	0.63	0.76	0.39
Random Forest	tvec	0.96	0.79	0.75	0.87
Naïve Bayes	tvec	0.92	0.84	0.87	0.82

BEST PERFORMING
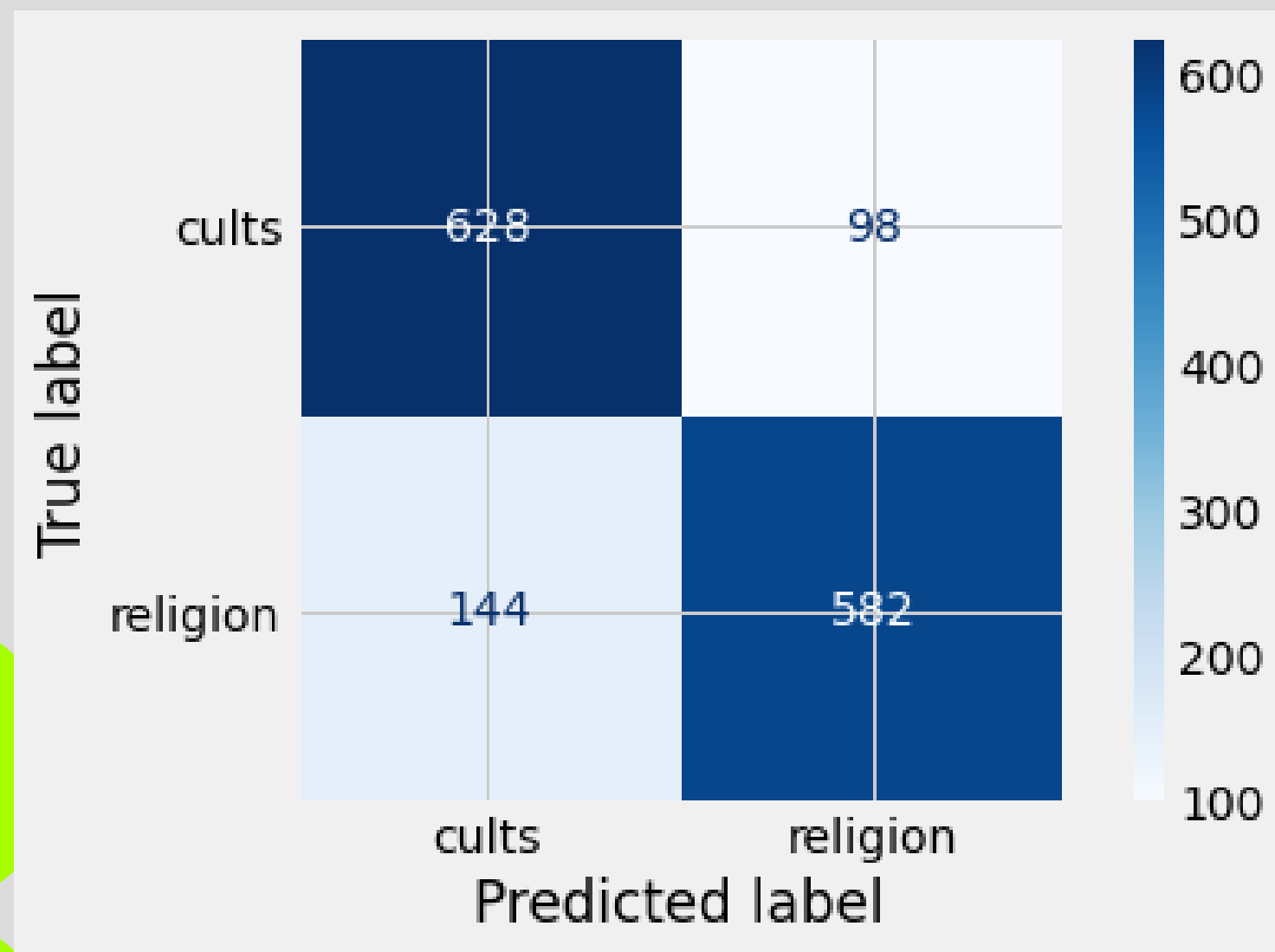
- training accuracy score **increased to 0.92** and the testing accuracy score increased compared to the Naïve Bayes baseline model to **0.83**.

- **Recall of cults – 87%:** We identified 87% of the actual text from the r/cults subreddit. This is about the same as the Naïve Bayes model in the baseline evaluation above.

- **The precision of cults: 82%:** When we predicted the text came from r/cults we got it right 81% of the time. So, 19% of our predictions inaccurately predicted r/cults. This is an improvement from the baseline.

MODEL EVALUATION

- A ROC Curve summarizes the trade-off between the true positive rate (tpr) and false positive rate (fpr) for a predictive model using different probability thresholds.

- Given that c is a constant known as decision threshold, Looking at the ROC curve to the right our blue curve is closer to the upper left corner than the 45-degree line
  - that suggests that this model performed quite well in being able to differentiate the two groups of religion and cults.

- Our ROC AUC is 0.92 this is a positive sign that our model performed well because the closer it is to one the better our model is at predicting with accuracy.


ROC Curve

MultinomialNB (AUC = 0.93)
baseline

IMPORTANT FEATURES

Most Important Features for predicting r/Religion

Most Important Features for predicting r/Cults

IMPORTANT FEATURES INSIGHTS

- The bar graph above shown in the previous slide is the top 10 most important features for predicting each of the respective subreddits.

- Commonalities:
  - **People**– Since both these areas are referring to a "social group" it is understandable that the term people would stand out.

  - **Church** – Interesting that both have this term. This could be attributed to the history of cults within religious practices.

  - **Think**– a concept that is controversial but present in dangerous groups is this idea of mind–control leaving the group unable to think for themselves.

- differences:
  - god/Jesus vs leader
  - members, join, group
  - question vs help
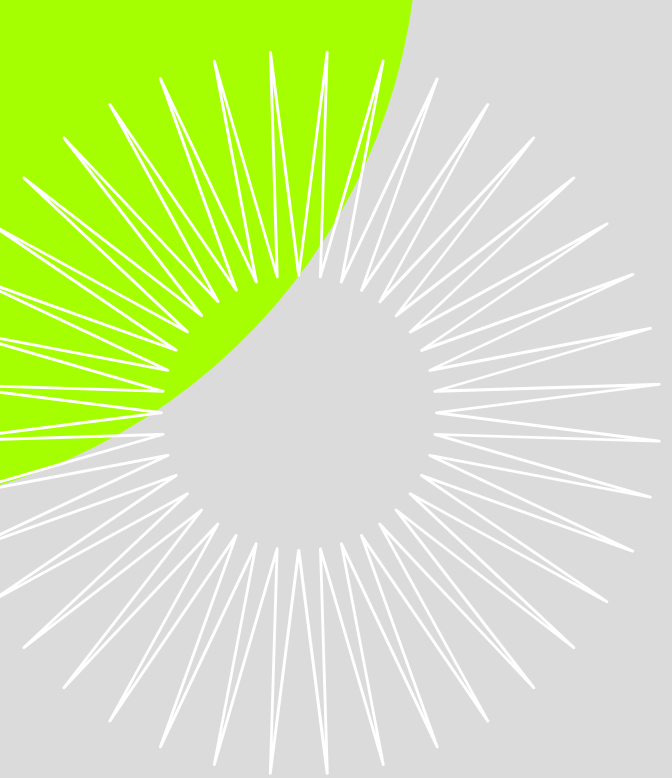
In order to develop better definitions for what defines a religion vs a cult, we looked to Reddit threads to understand how the language and communication styles differ within popular culture. Identifying the difference in vocabulary, and sentiment within each of these subreddit threads tells us there are several commonalities in terminology. These commonalities tell us that these groups are very similar and these commonalities could cause an issue when scaling this model. In addition, all models were overfitted telling us that to be able to scale this project we need better regularization techniques. The differentiating words are promising for our goal. Further analysis within types of speech could provide a clearer understanding.

Overall our modeling showed that while there are many similarities between the language/vocabulary used across the two groups that there are some defining words that make these two groups stand that can allow us to make a prediction with 0.84% accuracy on the subreddits. The ROC Score of 0.92 tells us that the probability of predicting the positive class (religion) correctly is 92%.

 To develop this project further it would be worth analyzing posts on specific social media pages dedicated to a specific cult and religion and compare the sentiment, vocabulary, and part-of-speech.

CONCLUSION

QUESTIONS