

EC News

Active Era



ather



Reddit



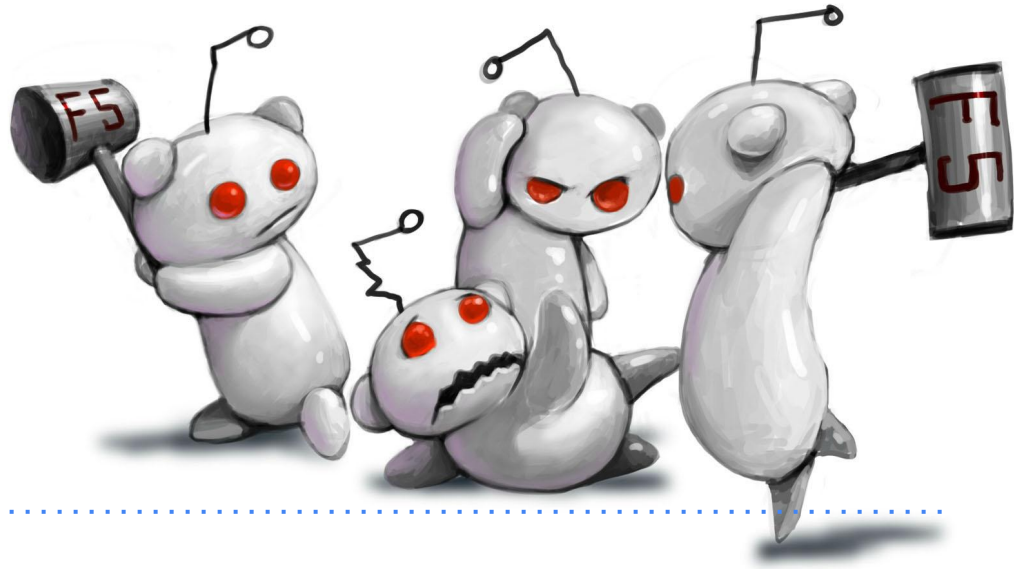
PureG

Aleisha Khan
GA Data Science Immersive - EC
Project 3

The Problem

Social media has been shown to promote and exacerbate polarization.

Is it possible to distinguish subreddits that foster open-minded dialogue from those that don't?



Subreddits

↑ 131 ↓  **r/unpopularopinion** · Posted by u/TheRealQwasson_ 18 hours ago

Walnuts in a brownie is an abomination to brownies

Brownies are fire. Walnuts are fine. But when you add walnuts to a brownie thats already rich, chocolatey, a bit gooey, and crisp on the outside it just destroys everything about it. I get mad whenever I eat a brownie and theres walnuts in it. They don't go well together, brownies should be eaten without walnuts. Why? Because they're soft and cakey, you add something solid like nuts then it immediately feels weird and out of place, its like adding raisins to a cookie. Why would you want raisins in a cookie??? Doesn't matter if the walnut is chopped or whole either, a brownie shouldn't have walnuts or any kind of nut. Ruins it. The only exception is peanut butter but that is butter from peanuts and it compliments the taste.

 23 Comments  Give Award  Share ...

85% Upvoted

↑ 15 ↓  **r/changemyview** · Posted by u/Snowman453 1 day ago

CMV: Capital punishment is bad

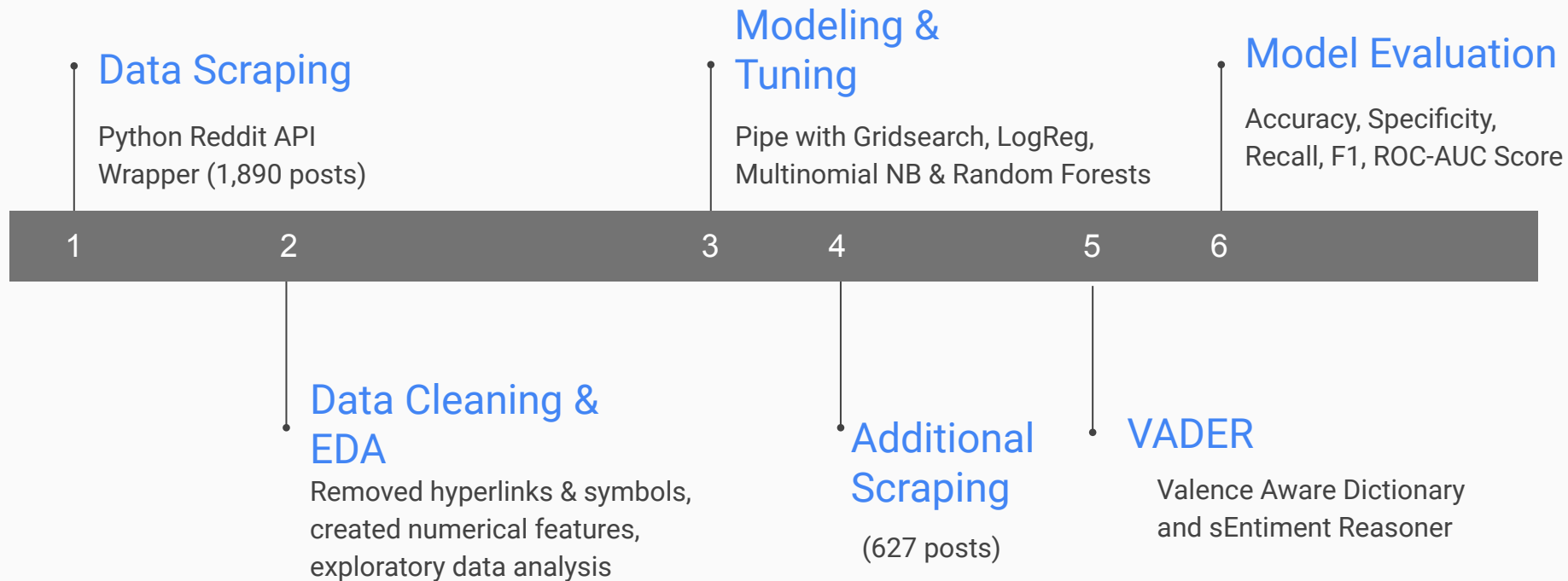
Delta(s) from OP

I personally believe that capital punishment is bad, but not for the reasons you may think. Most people, who receive capital punishment, commit evil acts, like murder. Should their lives end and not endure anymore pain, or should they suffer from the guilt they have to endure? Well, once someone dies, they no longer have to deal with pain anymore. They don't suffer from the guilt they have to endure. I personally think that those people are lucky and get a free pass. They should be suffering day and night until the day they die for the atrocities they committed. I don't understand why capital punishment was a thing. Please help me understand why. That would be greatly appreciated.

 32 Comments  Give Award  Share ...

80% Upvoted

The Process



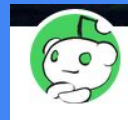
Scraping

Method

Python Reddit API Wrapper

Subreddits

Unpopular Opinion



Change My View



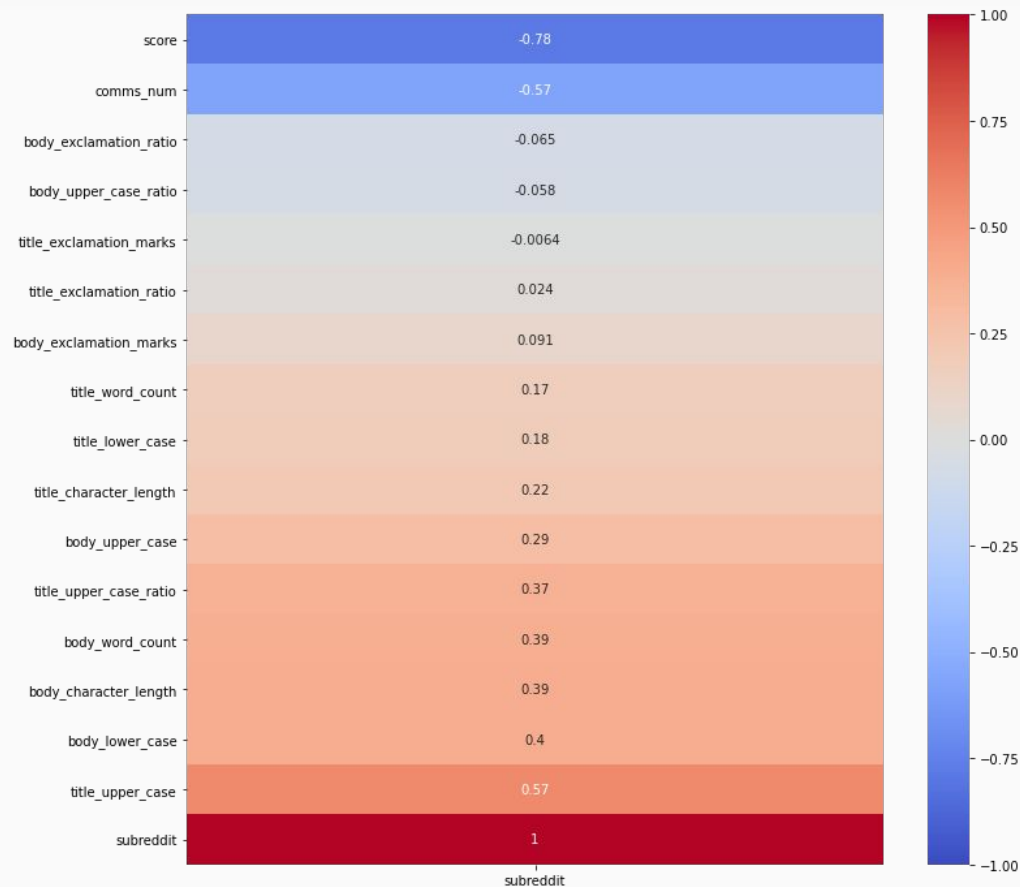
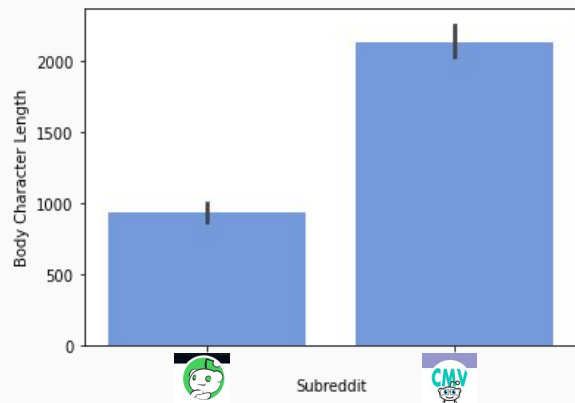
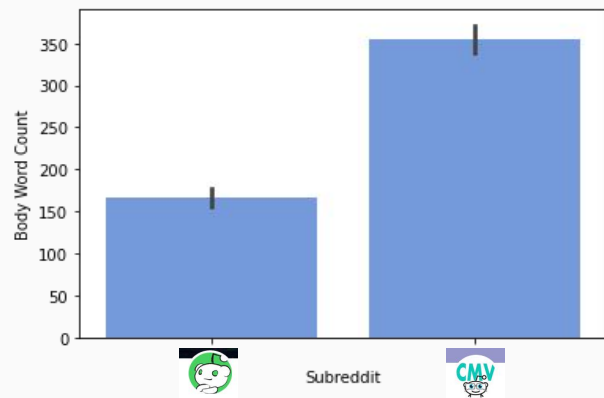
2,517 posts

05/2013 - 10/2020

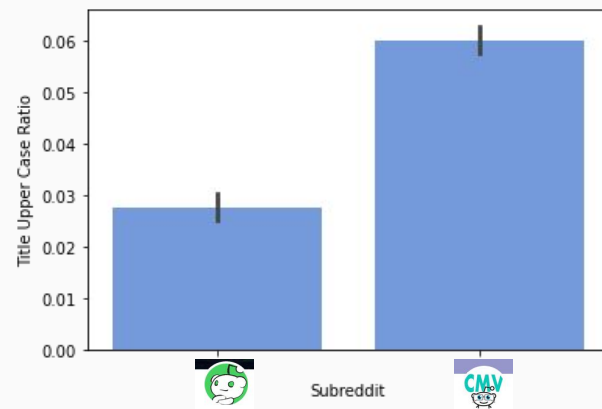
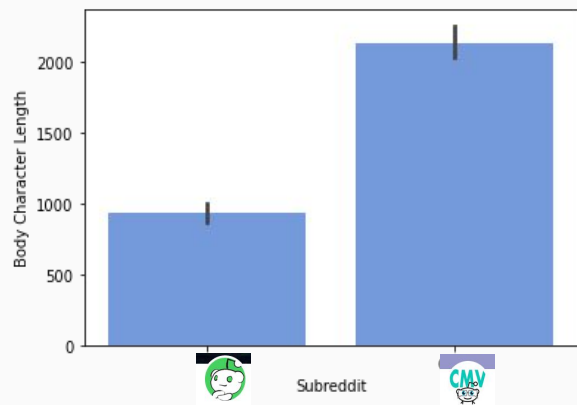
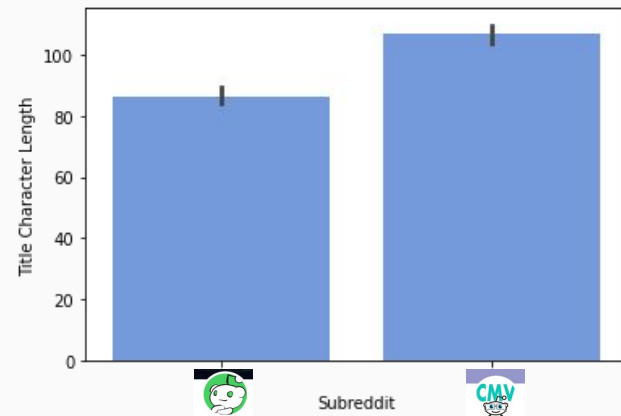
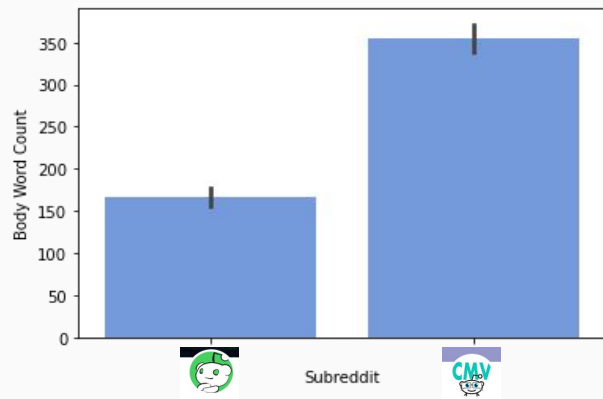
Data Cleaning & EDA

- 1.) Created numerical features
- 2.) Removed punctuation & hyperlinks (RegEx)
- 3.) Converted to all lowercase
- 4.) Identified the most common words
- 5.) Examined relationships between features and subreddit class

Numerical Features



Numerical Features

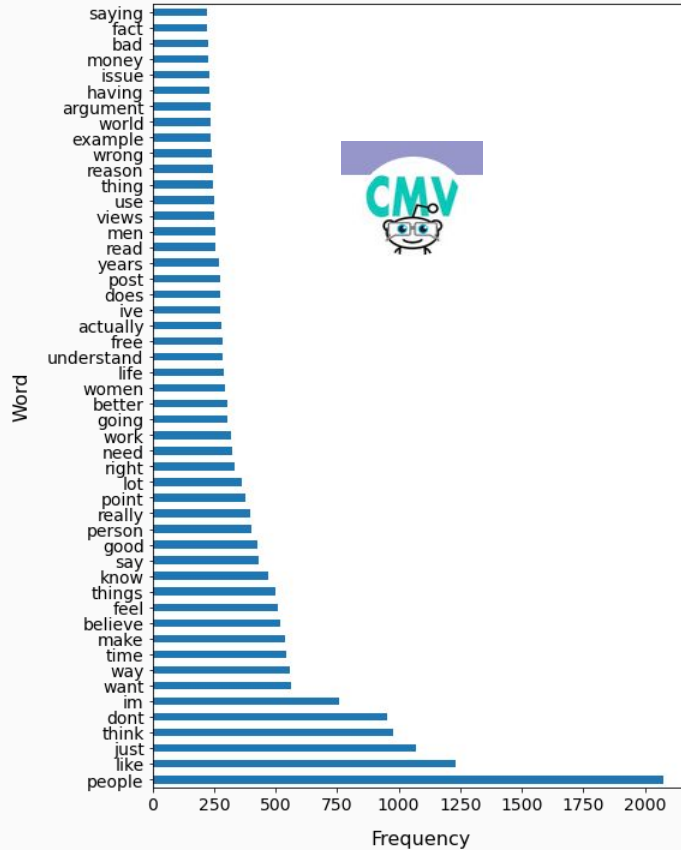


Word Frequencies

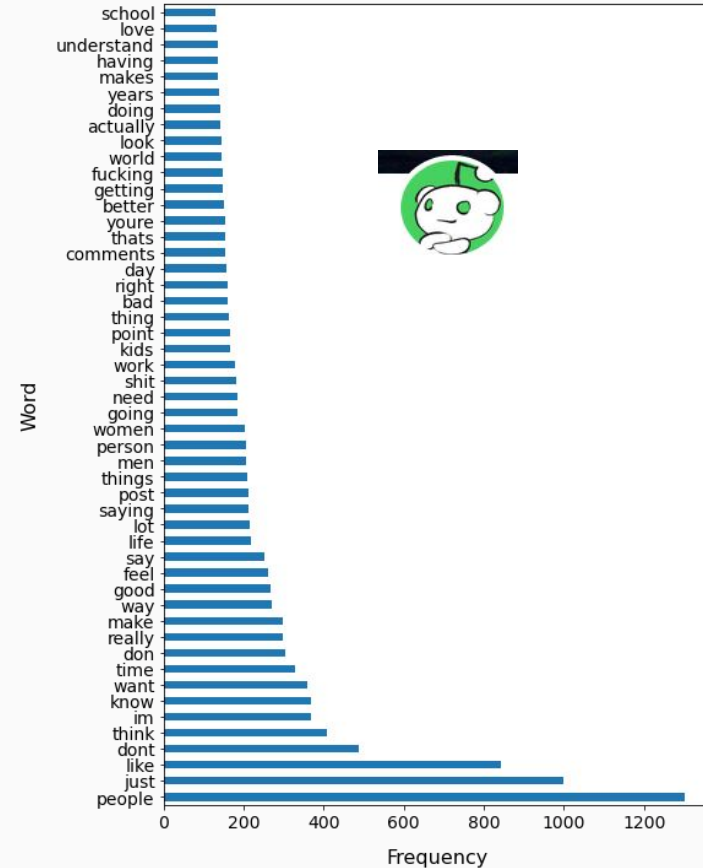


Word Frequencies

Frequency of Top 50 Words(ChangeMyView)



Frequency of Top 50 Words(UnpopularOpinion)



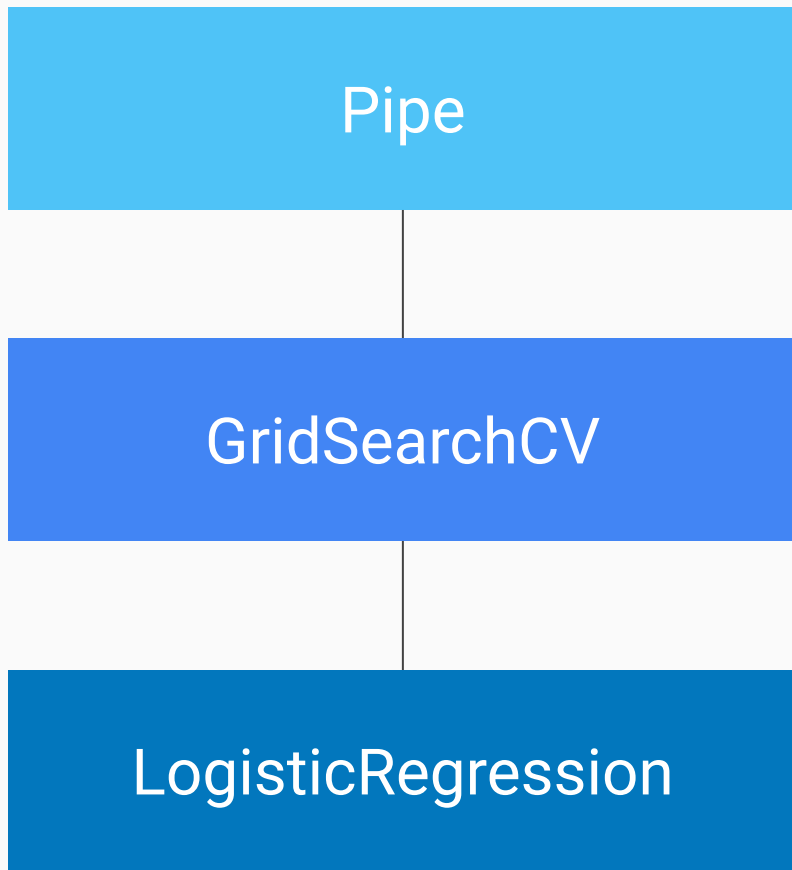
Modeling & Tuning

Original Baseline Accuracy : 57 %

Baseline Accuracy after Stratification: 42%

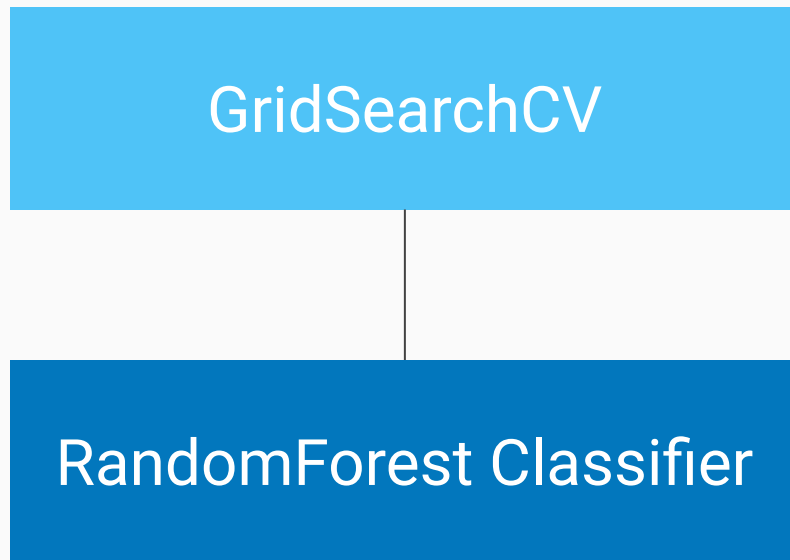
Model 1

Optimized parameters:
StandardScaler with_mean
and with_std,
LogisticRegression penalty,
solver, C, and max_iter.



Model 2

Optimized parameters:
n_estimators, max_depth,
and max_features.



Stopwords

Custom List

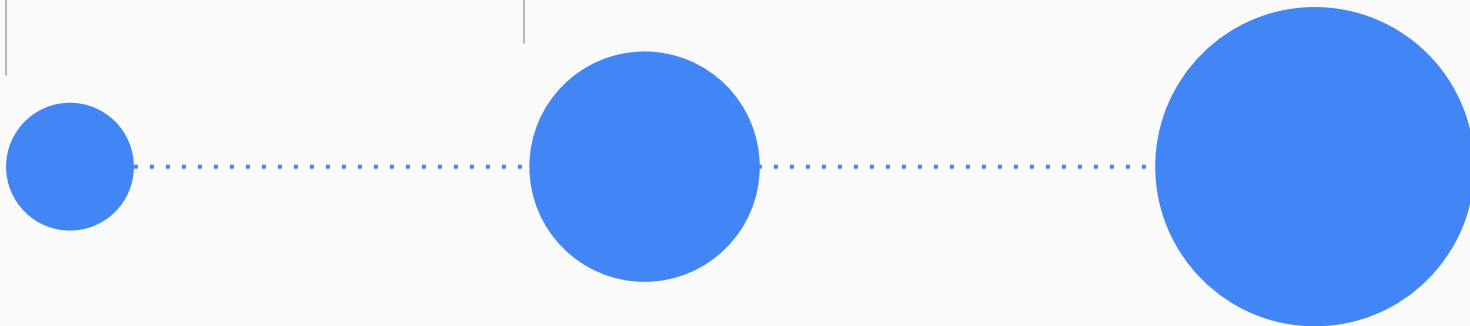
Cmv, change, view,
changemyview

NLTK English

NLTK's english
stopwords + my custom
list

Scikit-Learn English + Custom

Scikit-Learn's english
stopwords + my custom list



Model Evaluation

Model 1

Logistic Regression

Accuracy score 0.8531746031746031

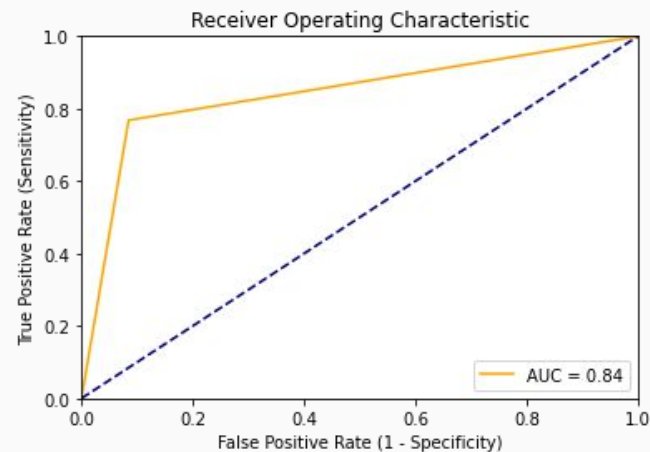
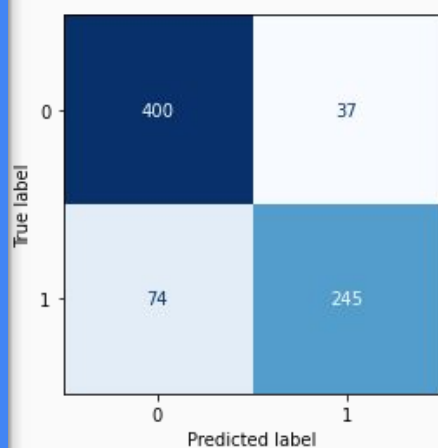
```
-----
                Predicted Unpopular Opinion  Predicted Change My View
Unpopular Opinion                400                37
Change My View                   74                245
-----
```

```
-----
                precision    recall  f1-score   support

     0         0.84         0.92         0.88         437
     1         0.87         0.77         0.82         319

 accuracy         0.85         0.85         0.85         756
 macro avg        0.86         0.84         0.85         756
 weighted avg     0.85         0.85         0.85         756
-----
```

ROC_AUC SCORE 0.9350372660559672



Model 2

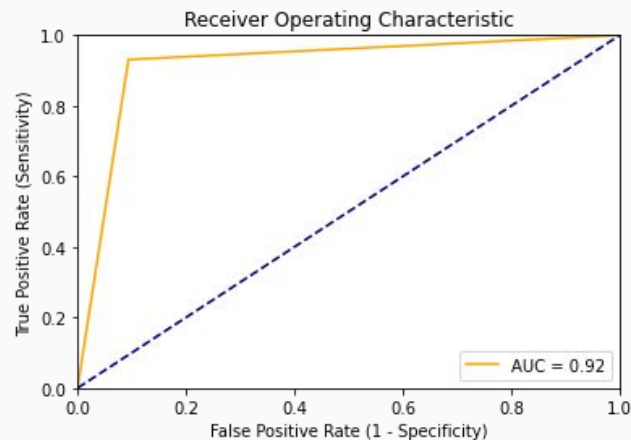
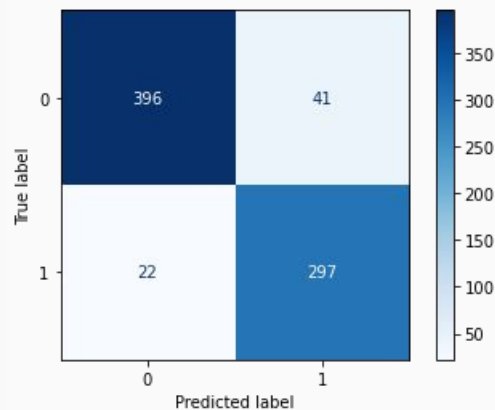
Random Forest Classifier

Accuracy score 0.9166666666666666

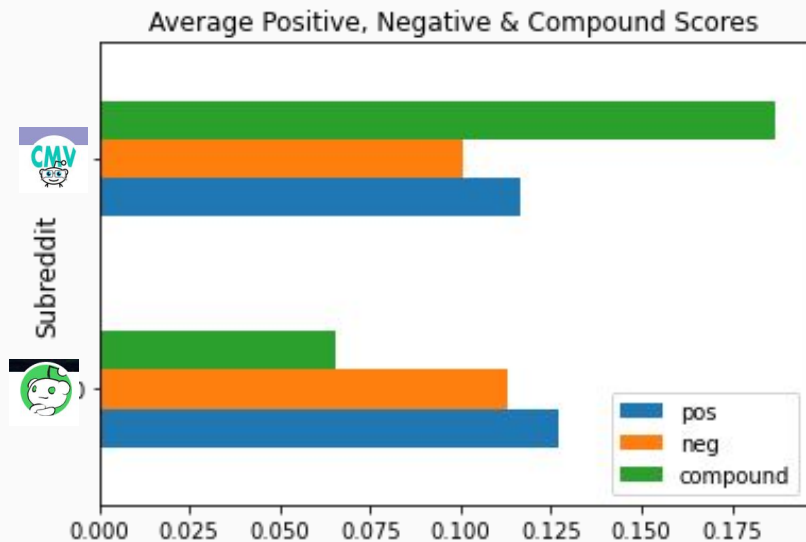
	Predicted Unpopular Opinion	Predicted Change My View
Unpopular Opinion	396	41
Change My View	22	297

	precision	recall	f1-score	support
0	0.95	0.91	0.93	437
1	0.88	0.93	0.90	319
accuracy			0.92	756
macro avg			0.91	756
weighted avg			0.92	756

ROC_AUC SCORE 0.9595848009009849



Valence Aware Dictionary and sEntiment Reasoner (VADER)



- Assigns a valence (positive, negative, neutral) using a rule-based dictionary.
- Does not pay attention to context
- Compound score tells us whether it is overall negative or positive

Model 1

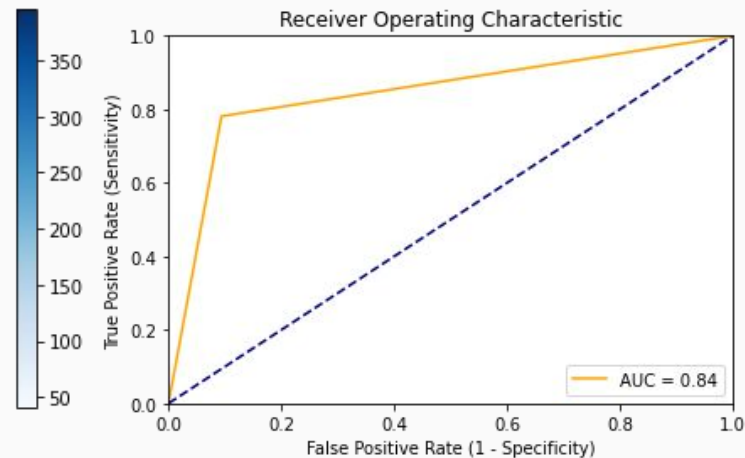
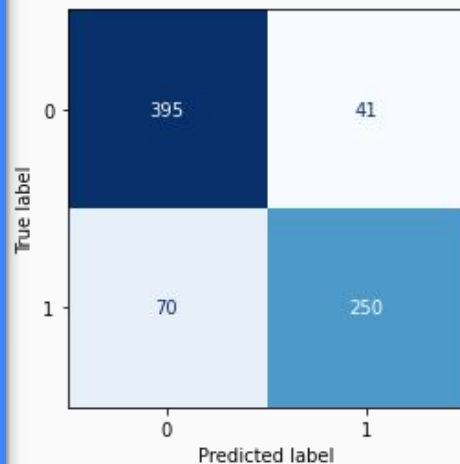
Logistic Regression w/
VADER

Accuracy score 0.8531746031746031

	Predicted Unpopular Opinion	Predicted Change My View
Unpopular Opinion	395	41
Change My View	70	250

	precision	recall	f1-score	support
0	0.85	0.91	0.88	436
1	0.86	0.78	0.82	320
accuracy			0.85	756
macro avg	0.85	0.84	0.85	756
weighted avg	0.85	0.85	0.85	756

ROC_AUC SCORE 0.9310636467889908



Model 2

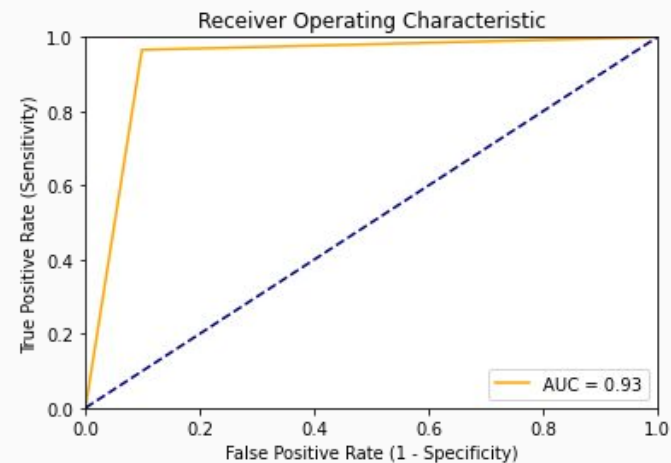
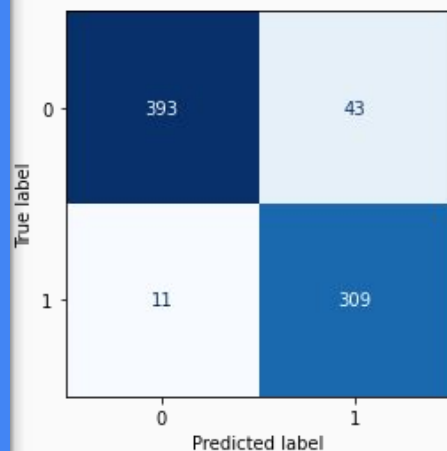
Random Forest Classifier w/
VADER

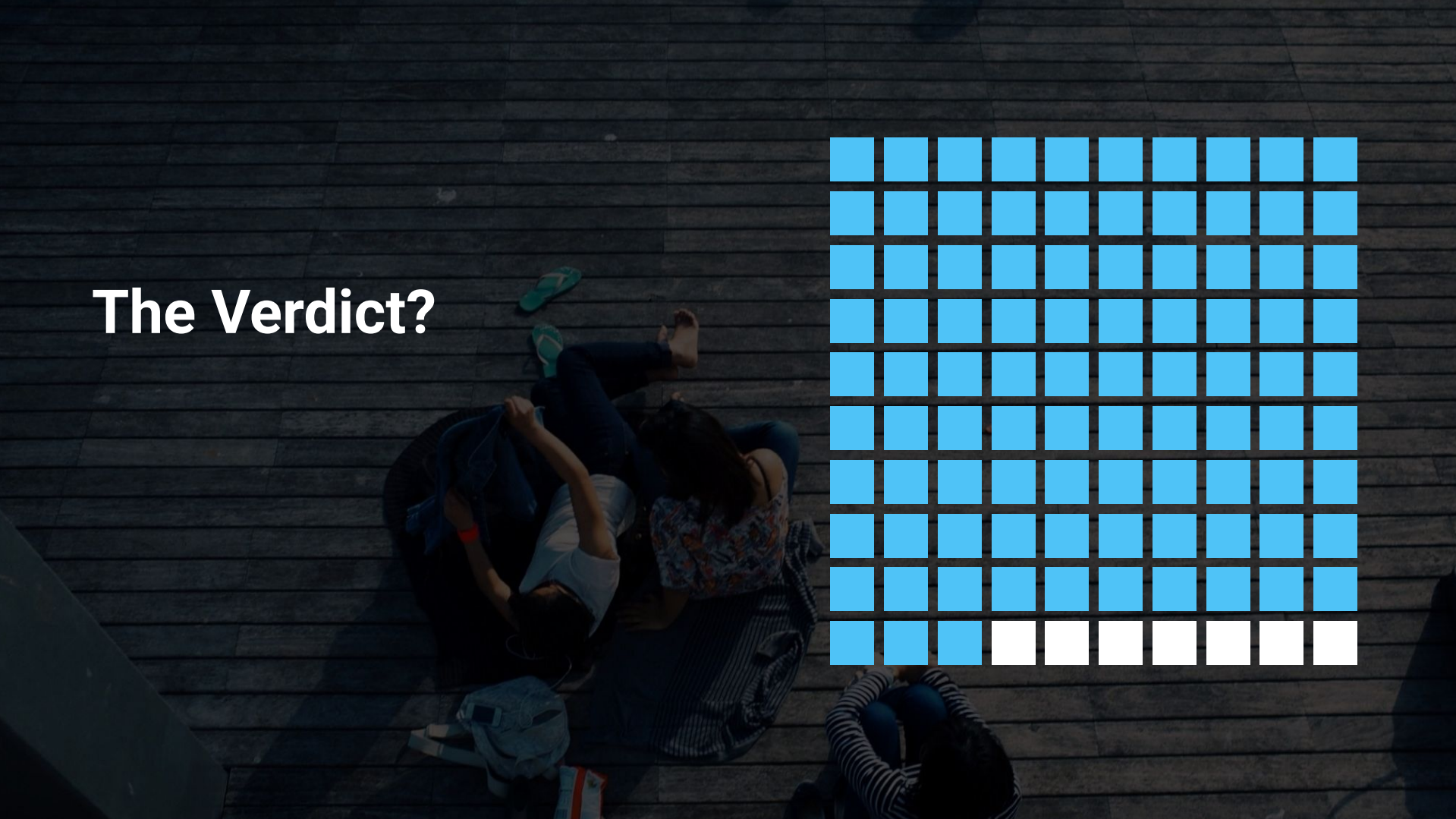
Accuracy score 0.9285714285714286

	Predicted Unpopular Opinion	Predicted Change My View
Unpopular Opinion	393	43
Change My View	11	309

	precision	recall	f1-score	support
0	0.97	0.90	0.94	436
1	0.88	0.97	0.92	320
accuracy			0.93	756
macro avg	0.93	0.93	0.93	756
weighted avg	0.93	0.93	0.93	756

ROC_AUC SCORE 0.9680225057339449



An aerial photograph of a group of people sitting on a wooden deck. A 10x10 grid of squares is overlaid on the right side of the image. The top 9 rows of the grid are blue, and the bottom row is white. The text 'The Verdict?' is written in white on the left side of the image.

The Verdict?