

The Daily News

Friday 10/08/2021

Moon Newsletter

Page 1

“Reddit Modeling Project”

-Thomaz Moon

- Is the World Getting Crazier?
- Is Satirical News Getting too Realistic?
- Turn to Page 2 to Find Out



Problem Statement

Friday 10/08/2021

Moon Newsletter

Page 2



Gullible or Warranted Beliefs?

Create a model to try to tell the difference between articles from r/TheOnion subreddit vs r/News subreddit.

Table of Content

Friday 10/08/2021

Moon Newsletter

Page 3

01

Modeling Process

How I went about testing models and parameters

02

Failures

What I tried and why it didn't work

03

Successful Models

What ended up working for me

04

Evaluations

Evaluating my models and metrics

05

Conclusions

Conclusions and recommendations

06

Game Time!

Can you identify what my model couldn't?

The Daily News

Friday 10/08/2021

Moon Newsletter

Page 4

A black and white photograph of a person's hands holding a folded newspaper. The person is wearing a dark jacket. The newspaper has some text and a small image visible on its pages.

01

Modeling Process

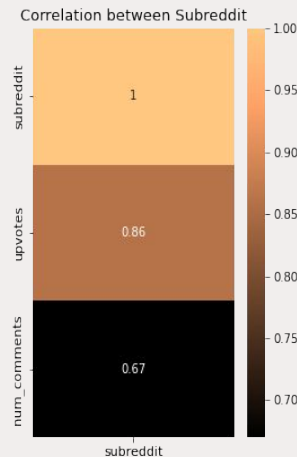
Modeling Process

Friday 10/08/2021

Moon Newsletter

Page 5

- Scraped 1k Posts From Each Subreddit w/ PRAW
- Basic Cleaning and EDA
- Fitting a Few Different Models and Fine Tuning Parameters
- Recording Responses
- Look for Possible Improvements

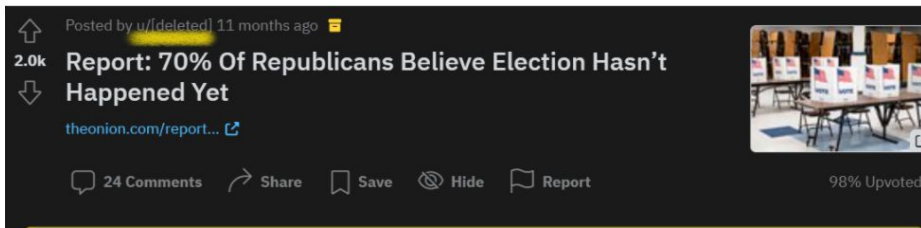


```
title          0
upvotes        0
num_comments   0
author         61
subreddit      0
dtype: int64
```

```
1 # Lets just get the first missing data and since this is on reddit, we can actually check instead of speculate
2 big_df[big_df['author'].isna()].iloc[0, 0]
```

'Report: 70% Of Republicans Believe Election Hasn't Happened Yet'

This author is na 7m assuming because the authors deleted their account.
it is



The Daily News

Friday 10/08/2021

Moon Newsletter

Page 6



02

What Didn't Work

Failures along the way

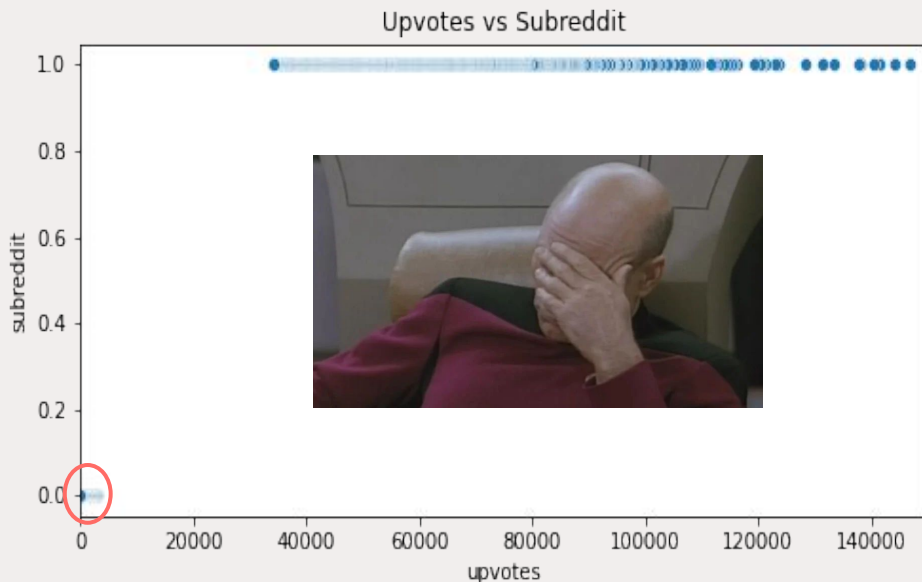
Friday 10/08/2021

Moon Newsletter

Page 7

Too Good to be True

- By using upvotes to determine subreddit, I made a perfect model
- Decided to check the number of upvotes by subreddit
- **TheOnion**: 164k members
News: 23.7 million members

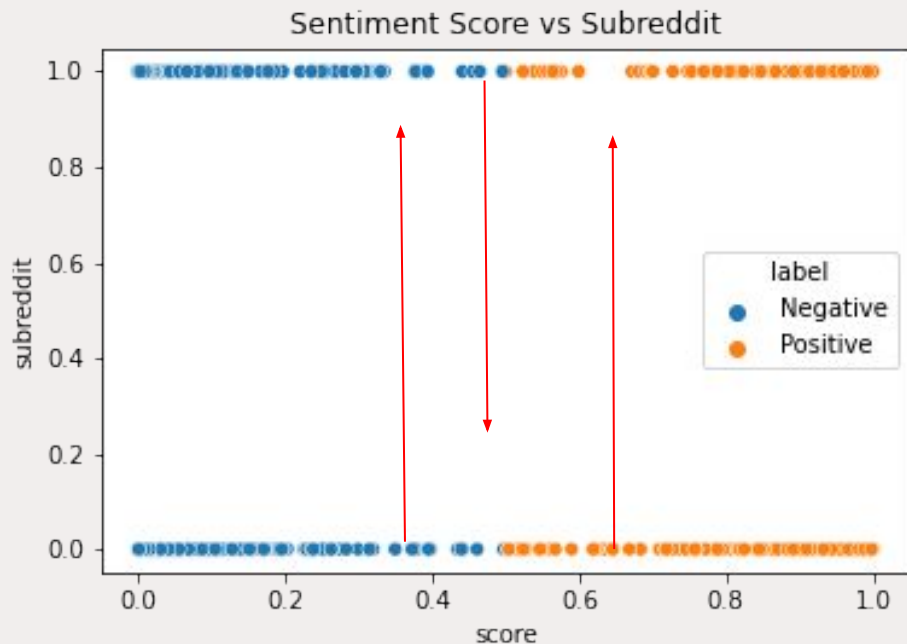


Failures along the way

Friday 10/08/2021

Moon Newsletter

Page 8



- Using Sentiment Analysis on the titles
- Logistic Regression scores were only around 56% and 55%
- Decided to check the distribution between subreddits

The Daily News

Friday 10/08/2021

Moon Newsletter

Page 9



Successful Models

Friday 10/08/2021

Moon Newsletter

Page 10

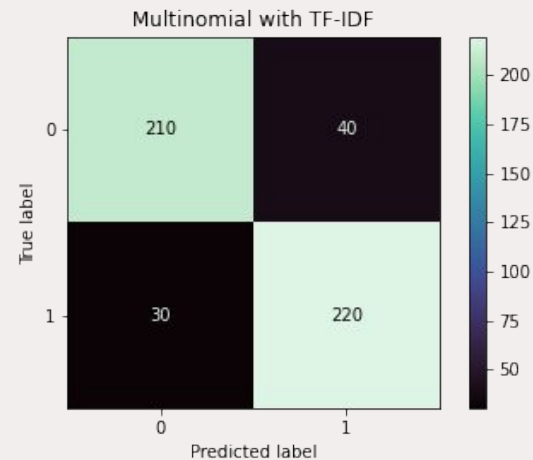
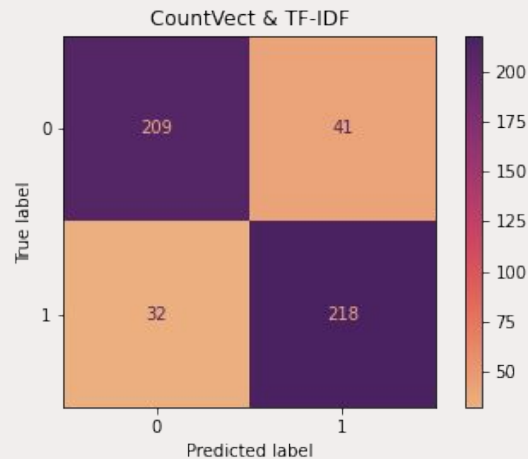
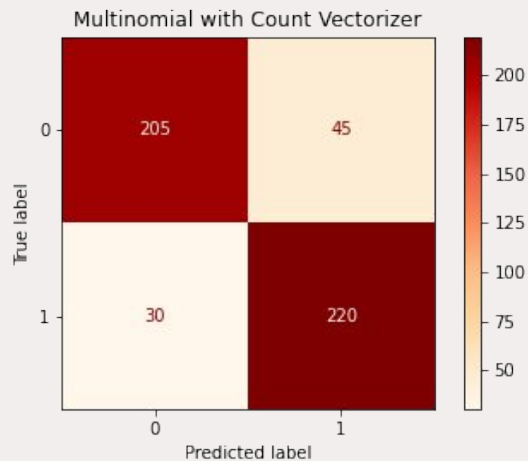
		training score	testing score	AUC	Misclassification Rate	Sensitivity	Specifcity	percision
23	mnb_gs	0.959947	0.850	0.926784	0.150	0.880	0.820	0.830189
	tfidf_gs	0.966622	0.860	0.92232	0.140	0.880	0.840	0.846154
	svc_gs	1.000000	0.850	0.921296	0.150	0.856	0.844	0.845850
	combo_gs	0.993324	0.836	0.91528	0.164	0.844	0.828	0.830709
	rf_gs	0.997330	0.824	0.903888	0.176	0.820	0.828	0.826613
1	Aggregated	0.986649	0.862	Not calculated	0.138	0.884	0.840	0.846743

Successful Models

Friday 10/08/2021

Moon Newsletter

Page 11

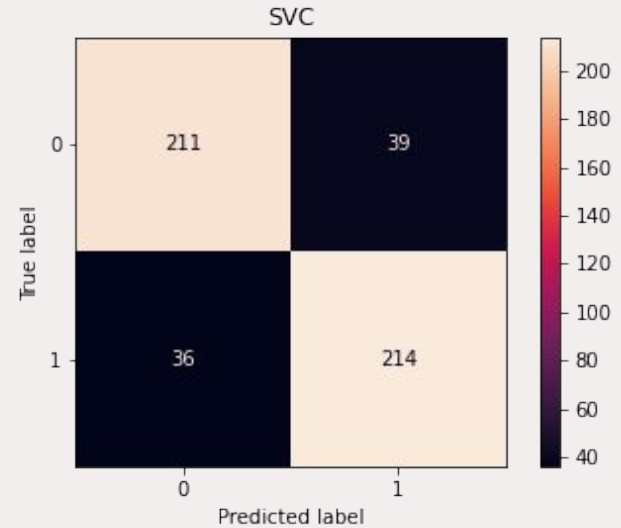
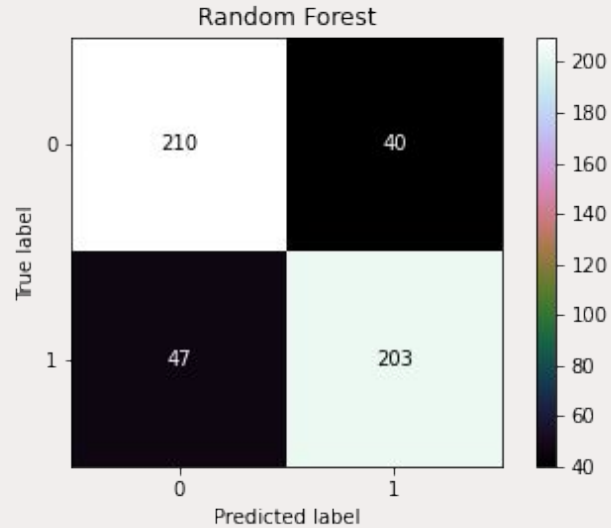


Successful Models

Friday 10/08/2021

Moon Newsletter

Page 12



The Daily News

Friday 10/08/2021

Moon Newsletter

Page 13

04

Evaluations



Goals

Friday 10/08/2021

Moon Newsletter

Page 14



- Beat the Baseline Accuracy of 50%
- Primarily Focus on Accuracy Over Other Metrics
- Have the Model Perform Well on New Unseen Data

How I did

Friday 10/08/2021

Moon Newsletter

Page 15

	training score	testing score	AUC	Misclassification Rate	Sensitivity	Specificity	percision
mnb_gs	0.959947	0.850	0.926784	0.150	0.880	0.820	0.830189
tfidf_gs	0.966622	0.860	0.92232	0.140	0.880	0.840	0.846154
svc_gs	1.000000	0.850	0.921296	0.150	0.856	0.844	0.845850
combo_gs	0.993324	0.836	0.91528	0.164	0.844	0.828	0.830709
rf_gs	0.997330	0.824	0.903888	0.176	0.820	0.828	0.826613
Aggregated	0.986649	0.862	Not calculated	0.138	0.884	0.840	0.846743

- All models had an accuracy of over 82%
- Fairly low misclassification rate
- A bit of variance, but with word vectors and unseen words in news can be expected

The Daily News

Friday 10/08/2021

Moon Newsletter

Page 16

05

Conclusions



Conclusions

Friday 10/08/2021

Moon Newsletter

Page 17

- While our models may not have been perfect, I think they're good enough to show that machines can tell the difference between satirical headlines and real news.
- Shows there's either:
 - some form of bias confirmation
 - Lack of research
 - Credulous personality

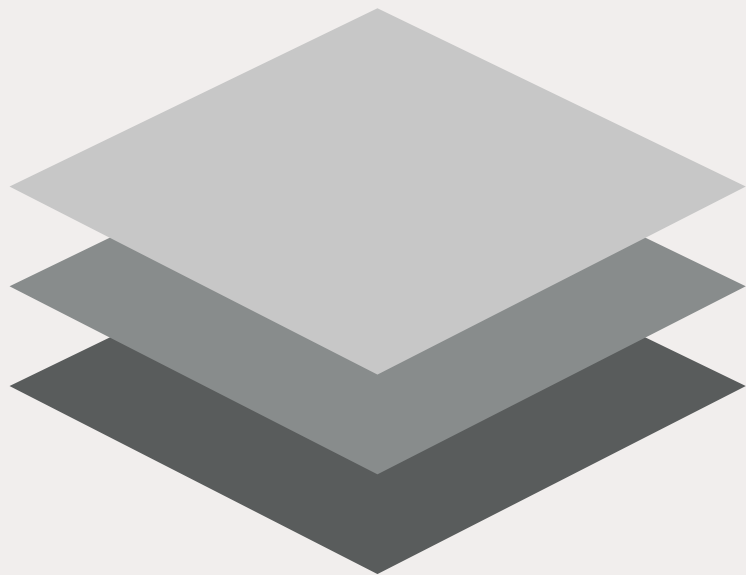


Recommendations

Friday 10/08/2021

Moon Newsletter

Page 18



Scrape more data in order to have more words when training



Make more models and aggregate their scores even more.



Start checking what websites the news articles are coming from

Are you better than a machine?

Friday 10/08/2021

Moon Special

Page 19

*Guess what subreddit these
headlines belong to!*



Guess the subreddit

Friday 10/08/2021

Moon Special

Page 20

1.

Supreme Court Waits In Line For Hours Before Voting To Uphold Arizona Restrictions

4.

9 juveniles injured in gunfight that broke out at 12-year-old's birthday party

2.

Police Now Ignoring Active Shooter Reports In Effort To Deprive Killers Of Attention

5.

12-Year-Old Boy Scouts Offer To Give Breast Exams

3.

Oklahoma mom of 11 helps rescue 10 girls on Afghanistan's robotics team

6.

Maryland Officers De-escalate Situation, Offer Compassion to Man in Behavioral Crisis

Guess the subreddit

Friday 10/08/2021

Moon Special

Page 21



Supreme Court Waits In Line For Hours Before Voting To Uphold Arizona Restrictions



9 juveniles injured in gunfight that broke out at 12-year-old's birthday party



Police Now Ignoring Active Shooter Reports In Effort To Deprive Killers Of Attention



12-Year-Old Boy Scouts Offer To Give Breast Exams



Oklahoma mom of 11 helps rescue 10 girls on Afghanistan's robotics team



Maryland Officers De-escalate Situation, Offer Compassion to Man in Behavioral Crisis

Sources

Friday 10/08/2021

Moon Newsletter

Page 22

- [Slide Template](#)

- Came with pictures I used

- [Angry kid](#)

- [Game Time](#)

- [r/TheOnion](#)

- [r/News](#)

- [News Logo](#)

- [The Onion Logo](#)

Special Thank to Jenn and Eddy

- **Jen**

- Helped fix my slides
 - Gave pointers and comments for improvement
- recommended the template website I'm using right now

- **Eddy**

- Helped me with problem statements.
 - I ended up changing the subreddit, but his original ideas were fire

Thanks

Monday 01/02/2020

Newsletter

Edition: 001



Do you have any questions?
addyouremail@freepik.com
+91 620 421 838
yourcompany.com



CREDITS: This presentation template was created by **Slidesgo**, including icons by **Flaticon**, and infographics & images by **Freepik**

Please keep this slide for attribution