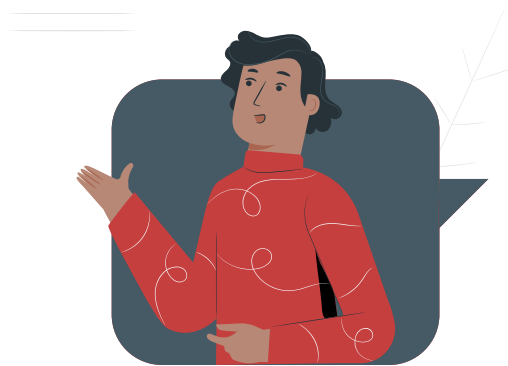


News or “News”?:

**Distinguishing facts
from misinformation
on  reddit**

Who are **we**?



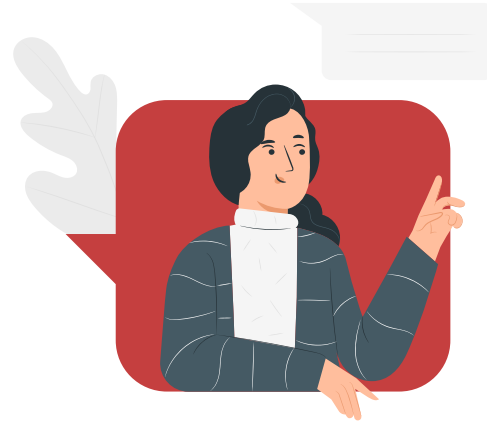
Arcavia Data Solutions

Who are **our stakeholders**?



Reddit

Primary Stakeholder



User Moderators

Secondary Stakeholder

Agenda

01

Context &
Problem Statement

02

Data Collection

03

Data Cleaning &
Exploratory Data Analysis

04

Model Evaluation

05

Implementation

06

Conclusion &
Recommendations

01

Context & Problem Statement

Congratulations!!!



Reddit

Reddit shares soar on first day of public trading

Company stock price climbed nearly 70% from its IPO, with Condé Nast's parent company standing to make nearly \$1.4bn

Kari Paul and agencies

Fri 22 Mar 2024 01:28 GMT

THE BUSINESS TIMES



Reddit prices IPO at top of indicated range to raise US\$748 million

Published Thu, Mar 21, 2024 · 7:01 am



The IPO priced 22 million shares sold by Reddit and its existing shareholders at US\$34 a share, giving the company a valuation of about US\$6.4 billion. PHOTO: REUTERS

However...



“Many of the anonymous political hobbyists and social hackers who are creating and **disseminating fake news** do so on **Reddit**...

However...



With around 300 million monthly users globally, **Reddit is the fourth most visited website in the US**, and eighth in the world, as of this month.”

Steven Rosenbaum
Contributor, Project Syndicate

Context



Reddit is sometimes referred to as the **“least known but most influential”**^[1] social media site with a significant ability to influence and shape public perception.

With a small team of in-house moderators, Reddit **relies heavily on user moderators** who volunteer to moderate content on the individual subreddits they belong to. As expected, such a decentralised and **“loose approach to content moderation”**^[2] has its pitfalls.

Sources:

[1] [Knowledge at Wharton, 2019](#) [2] [CNA, 2024](#)

Terminology



Misinformation is defined as “any information that is **demonstrably false** or otherwise **misleading**, regardless of its source or **intention**.”

– American Psychological Association

**Can you identify
misinformation?**



r/news



Search in r/news



r/news · 10 hr. ago

ILovePublicLibraries



Worcester library fees? No problem. Just show them your cat photos.



boston.com

Open

FACT

MISINFORMATION



r/TheOnion



Search in r/TheOnion



r/TheOnion • 7 days ago

dwaxe



Airbnb To Ban Indoor Security Cameras



theonion.com

Open

FACT

MISINFORMATION



r/news • 2 days ago

JussiesTunaSub



Teacher who resigned after her OnlyFans page was discovered says new employer fired her for violating social media policy



cleveland19.com

Open

FACT

MISINFORMATION

Who is Daniel?

Daniel is an **elementary school teacher**. With an **interest in current affairs and politics**, he spends his free time as a **volunteer moderator** on the **r/politics subreddit**.

What are his goals?

Daniel hopes to maintain the r/politics subreddit as a **safe and conducive environment** for civil, political discussions.



Daniel, 25

What does he believe in?

Daniel believes in **freedom of speech** and wants to **encourage more Americans to exercise their right to vote**.

What's affecting him recently?

With the upcoming **2024 US presidential elections**, Daniel is concerned about the **rise in online misinformation** and how this might impact the elections.

Problem Statement

As Reddit navigates its IPO lock-up this year, it has been attempting to adopt a **more serious approach** towards content moderation.

By using a classification model, we aim to help Reddit **reduce the spread of misinformation by determining whether news headlines on the site may be classified as fact or misinformation based on user engagement and discussion.**

Additionally, we aim to help the user moderators on Reddit **streamline their content moderation process** within their individual subreddits.

Sources



“News” Source

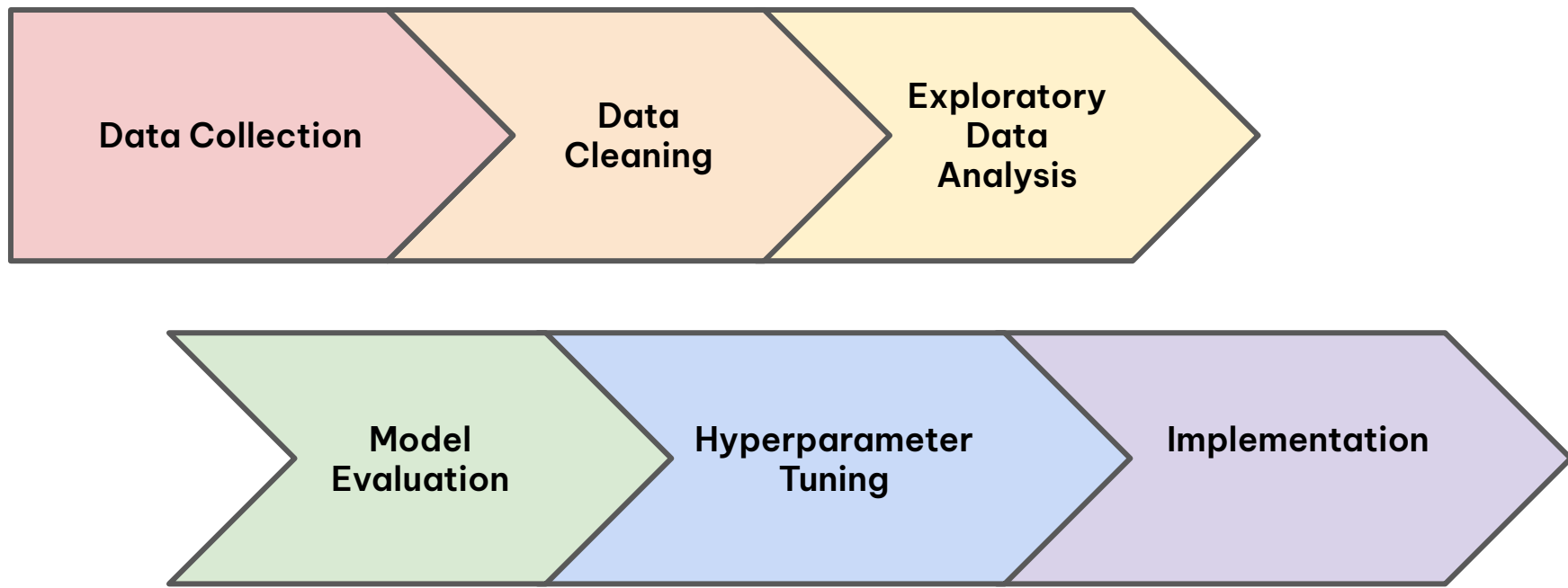


The Onion is the world's leading news publication, offering highly acclaimed, universally revered coverage of breaking national, international, and local news events. Rising from its humble beginnings as a print newspaper in 1756, *The Onion* now enjoys a daily readership of 4.3 trillion and has grown into the single most powerful and influential organization in human history.

In addition to maintaining a towering standard of excellence to which the rest of the industry aspires, *The Onion* supports more than 350,000 full- and part-time journalism jobs in its numerous news bureaus and manual labor camps stationed around the world, and members of its editorial board have served with distinction in an advisory capacity for such nations as China, Syria, Somalia, and the former Soviet Union. On top of its journalistic pursuits, *The Onion* also owns and operates the majority of the world's transoceanic shipping lanes, stands on the nation's leading edge on matters of deforestation and strip mining, and proudly conducts tests on millions of animals daily.

The Onion is now available exclusively online without charge in order to take advantage of various charity tax benefits.



The Data Science Process



02

Data Collection

Web Scrapping

	r/News	r/TheOnion
	 r/news	 r/TheOnion
No. of posts	100	100
No. of comments	19,985	17,934
	Facts	Misinformation

Sources:

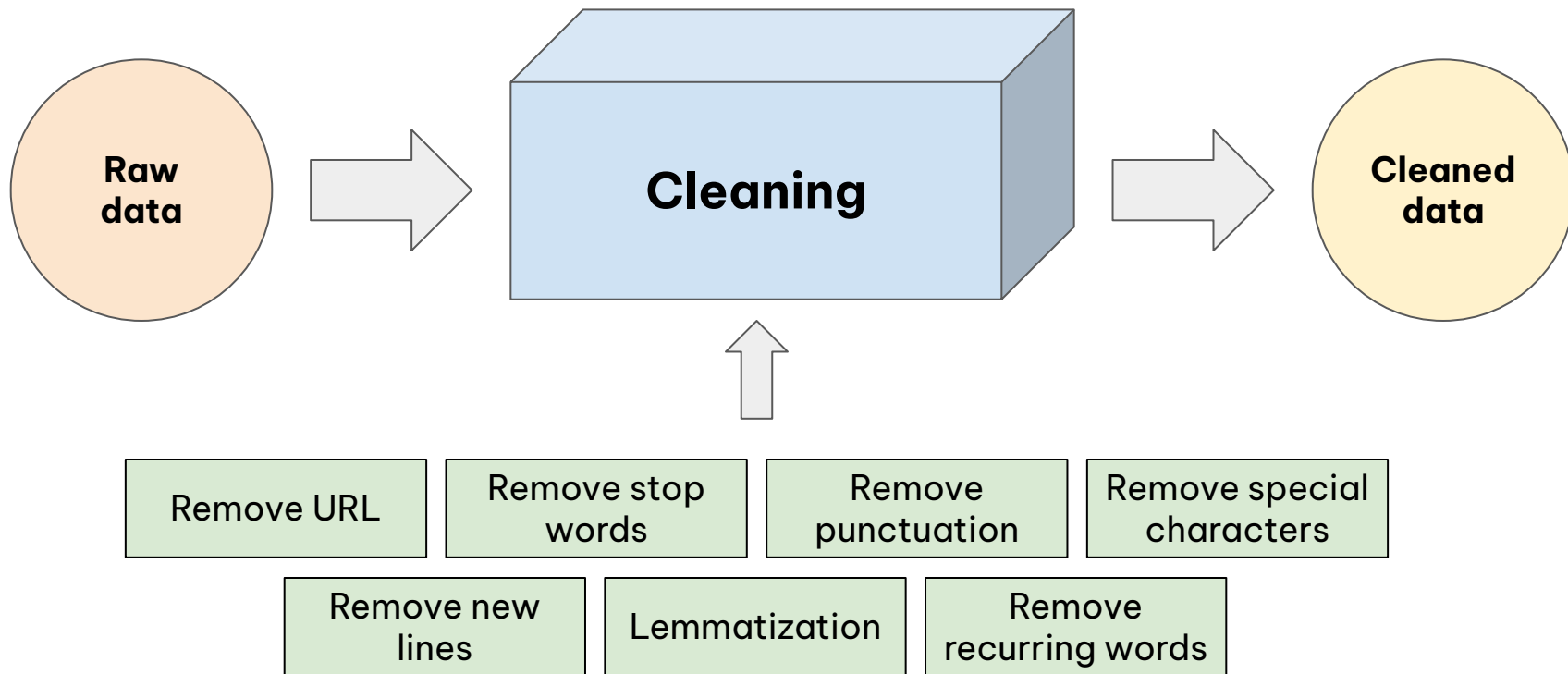
<https://www.reddit.com/r/news/>

<https://www.reddit.com/r/TheOnion/>

03

Data Cleaning & Exploratory Data Analysis

Data Cleaning



EDA

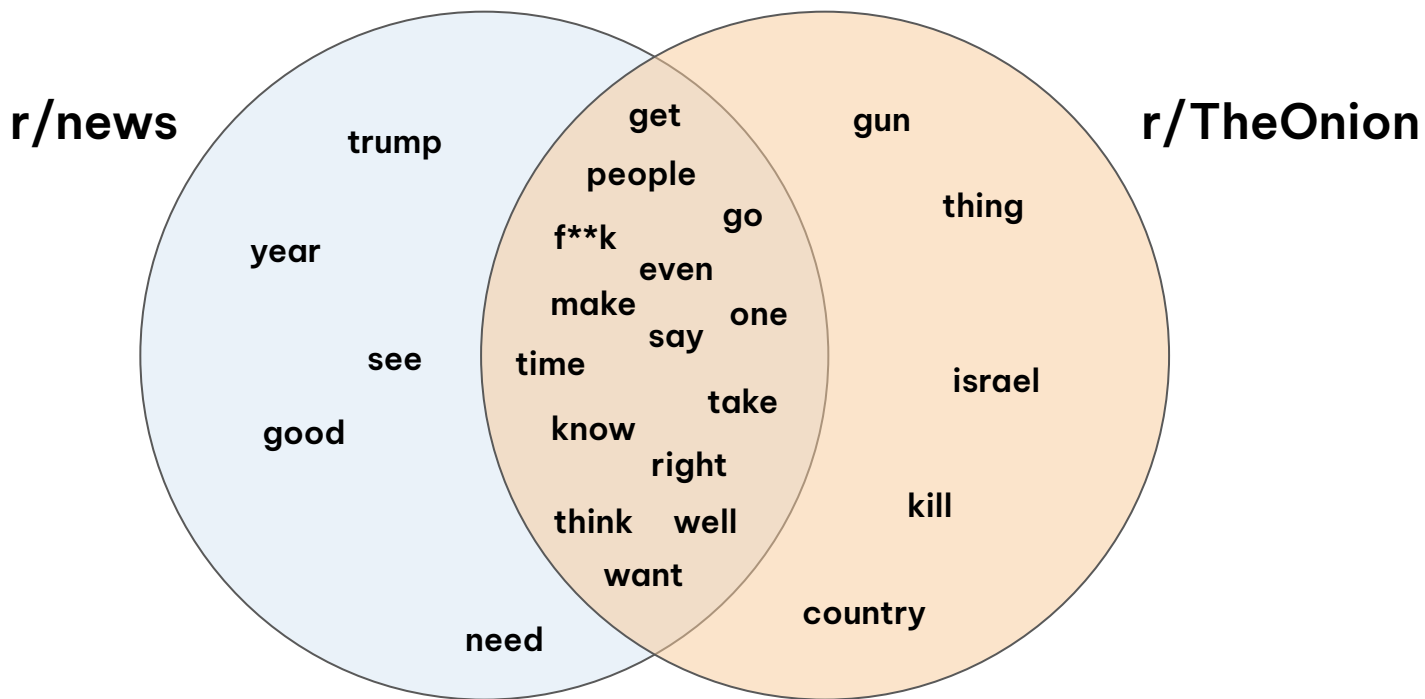
Word Cloud for r/news



Word Cloud for r/TheOnion

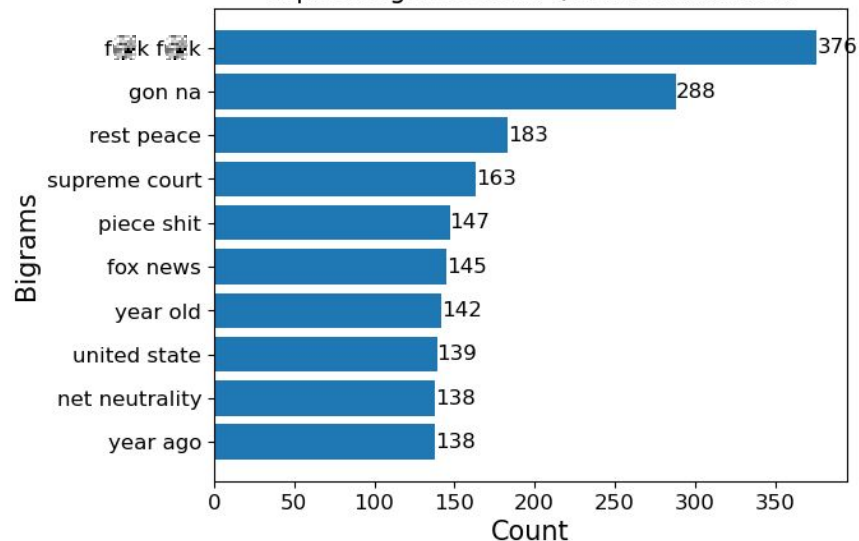


Common Top 20 Words

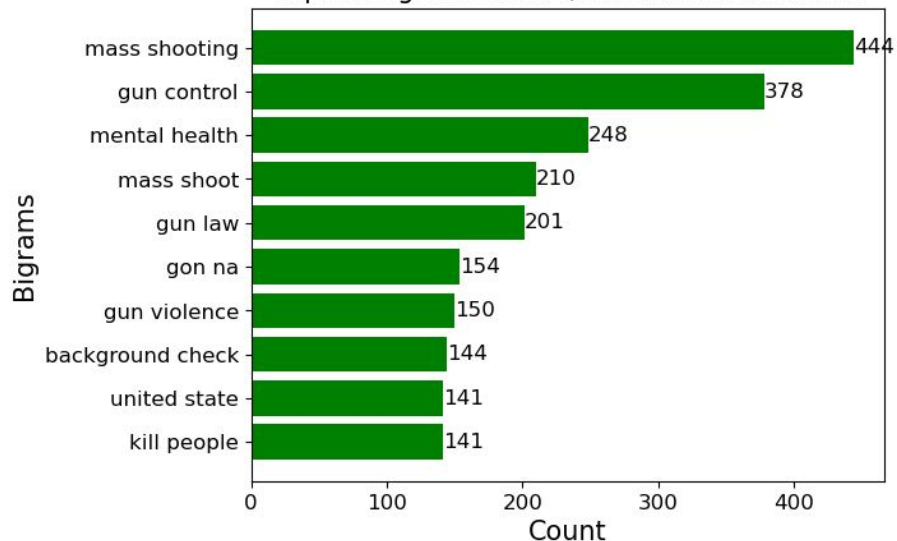


Top 10 Bigrams from Each Subreddit

Top 10 Bigrams from r/news Comments



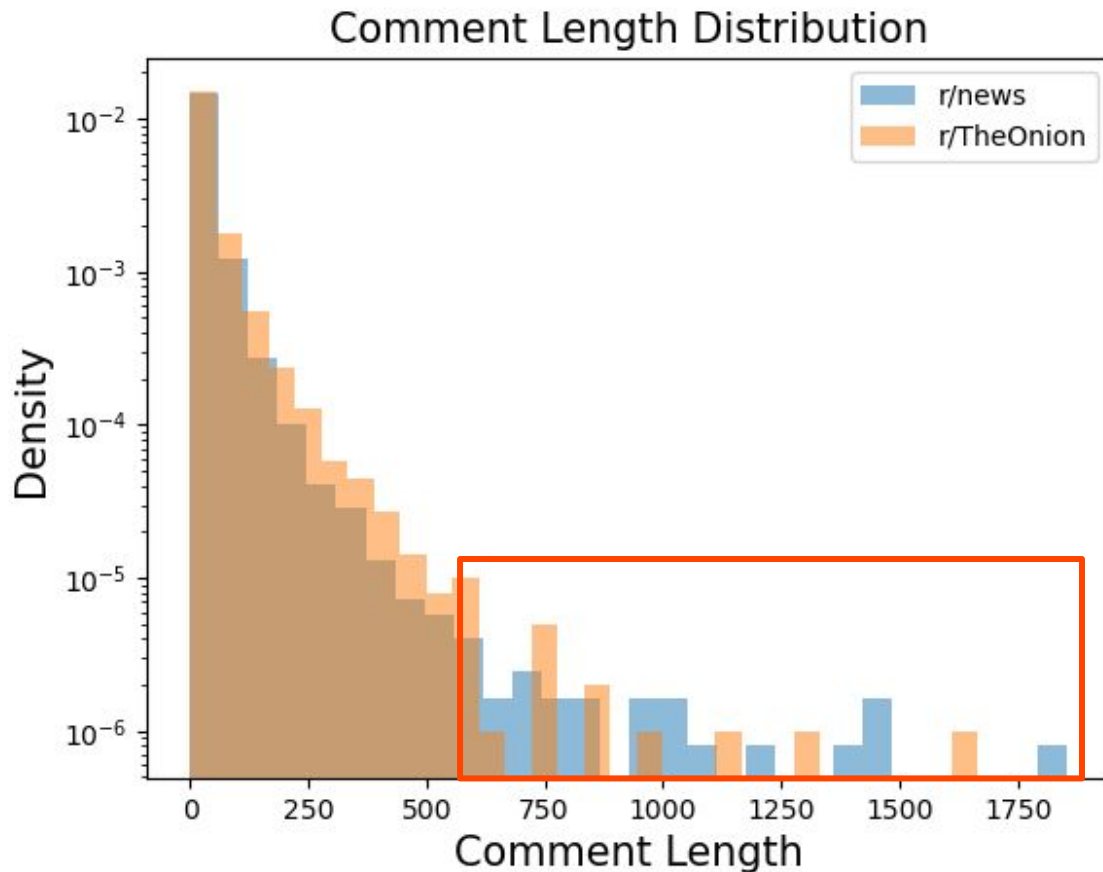
Top 10 Bigrams from r/TheOnion Comments



EDA

Comments from
r/TheOnion are
generally shorter
than those from
r/news.

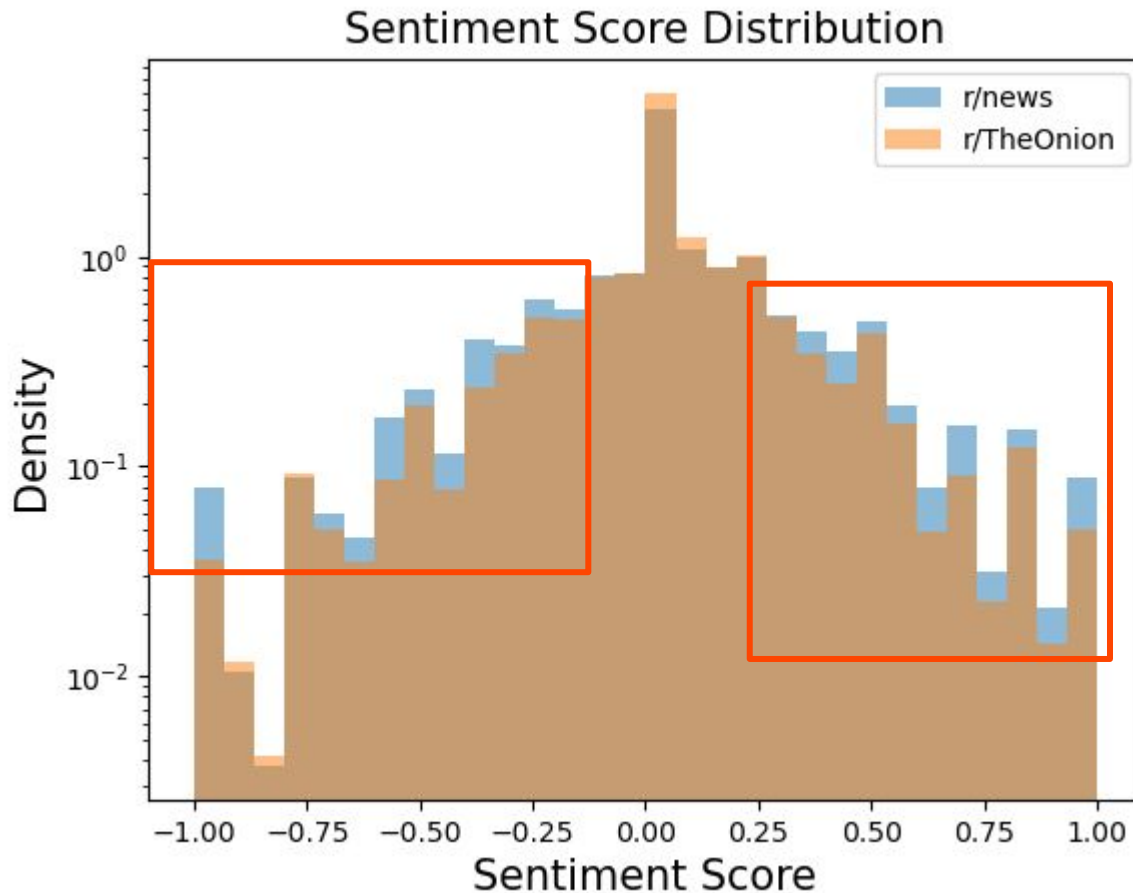
Comment Length Distribution



EDA

r/news comments tend to be **more polarised** (very positive and very negative).

Sentiment Score Distribution



04

Model Evaluation

Metrics

Accuracy

Indicates overall correctness of predictions

Sensitivity

Maximises correct identification of misinformation among all instances of actual misinformation

Precision

Maximises correct identification of misinformation among all instances classified as misinformation

F1 Score

Balances sensitivity and precision (useful for unbalanced classes)

Metrics

Maximise

Accuracy

Indicates overall correctness of predictions

Precision

Maximises correct identification of misinformation among all instances classified as misinformation



Sensitivity

Maximises correct identification of misinformation among all instances of actual misinformation



F1 Score

Balances sensitivity and precision (useful for unbalanced classes)

Model Evaluation

Model	k-Nearest Neighbors (kNN)	Naive Bayes	Logistic Regression	Bagging Classifier
Accuracy	52.96%	77.00%	76.88%	71.59%
Precision	50.05%	78.02%	74.23%	68.78%
Sensitivity	92.87%	70.87%	78.08%	72.73%
F1 Score	65.04%	74.14%	76.04%	70.67%

Model Evaluation

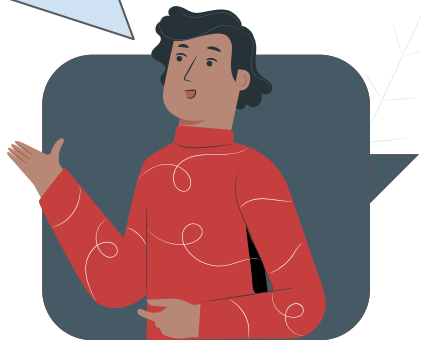
Model	k-Nearest Neighbors (kNN)	Naive Bayes	Logistic Regression	Bagging Classifier
Accuracy	52.96%	77.00%	76.88%	71.59%
Precision	50.05%	78.02%	74.23%	68.78%
Sensitivity	92.87%	70.87%	78.08%	72.73%
F1 Score	65.04%	74.14%	76.04%	70.67%

Hyperparameter Tuning





One of the hyperparameters we adjusted was `class_weights`.

Setting `class_weights` to “balanced” addresses the issue of class imbalance by **assigning higher weights to the instances of the minority class** during the training process.

Hyperparameter tuning is the process of finding the best settings for the machine learning model. **Different combinations of hyperparameters** are adjusted to get the **best performance** from the model.



Hyperparameter Tuning

	Before Tuning	After Tuning	
Model	Logistic Regression	Logistic Regression	
Accuracy	76.88%	80.02%	 +3.14%
Precision	74.23%	75.84%	 +2.17%
Sensitivity	78.08%	84.51%	 +6.43%
F1 Score	76.04%	79.94%	 +3.90%

05

Implementation

Background

 **NBC NEWS**

Disinformation poses an unprecedented threat in 2024 – and the U.S. is less ready than ever

The U.S. presidential election comes at a time of ideal circumstances for disinformation and the people who spread it.



Source:
[NBC News, Jan 2024](#)

Background

CNET

News > Politics

Reddit was a misinformation hotspot in 2016 election, study says

Researchers find a surge in links to controversial sites and redditors from "fringe" forums. The company's CEO says Reddit is "digging deeply" on potential abuse.

Meanwhile, [generative artificial intelligence tools](#) have made it far cheaper and easier to spread the kind of [misinformation that can mislead voters and potentially influence elections](#). And social media companies that once invested heavily in correcting the record have [shifted their priorities](#).

Sources:
[CNET, Dec 2017](#) and [AP News, Dec 2023](#)

Background



♥ DONATE

TECHNOLOGY

Managing Misinformation On Reddit

DECEMBER 8, 2019 · 8:01 AM ET

With 5.6 million members, Reddit's r/politics is a massive online forum. NPR's Lulu Garcia-Navarro **speaks with moderator "QuInlan" about managing discussions in an era of misinformation.**

Source:

[NPR, Dec 2019](#)

Reporter

GARCIA-NAVARRO: We here at NPR have been covering, quote, unquote, "fake news" in its original sense - you know, disinformation, misleading information masked as journalism or facts - for years now. How do you and your team of moderators deal with that?

**r/politics
Moderator**

QUINLAN: In short, we don't. We are a team of about 50 moderators on a subreddit of 5.6 million people. We don't have resources that a team like Facebook or Twitter might have with tons and tons of paid fact-checkers. So generally, when we see fake news that we know is not factually correct, we'll go ahead and encourage our user base to go ahead and fact-check it themselves and also comment - say, well, here's why this news is fake, and here is the sources to back it up. And here's why you shouldn't believe everything that you see on the Internet.

Sample Use Case



r/politics

8.5m members ● 5.9k online



It's been a long day at work...
I should probably look into
the r/politics subreddit but
there are too many new posts
from today and I don't know
where to start!

Daniel, 25

Elementary School Teacher
Reddit User Moderator

Sample Use Case: Misinformation Checker



r/politics

8.5m members ● 5.9k online



Select	Title
<input type="checkbox"/>	Biden Gets Some Good News in Poll as He Gains Ground on Trump
<input type="checkbox"/>	Biden calls for 'immediate ceasefire' in Gaza
<input type="checkbox"/>	Eric Trump says \$454m fine imposed on his father 'doesn't exist in this country'
<input type="checkbox"/>	RFK Jr. Apologizes for Super Bowl Ad ... That's Still Pinned Atop His Twitter Page
<input type="checkbox"/>	It's a date: Trump's first felony trial will be April 15
<input type="checkbox"/>	Joe Biden Pledges \$1.7 Billion To End Hunger Across U.S.

Sample Use Case: Misinformation Checker



r/politics

8.5m members ● 5.9k online



Select	Title
<input checked="" type="checkbox"/>	Biden Gets Some Good News in Poll as He Gains Ground on Trump
<input checked="" type="checkbox"/>	Biden calls for 'immediate ceasefire' in Gaza
<input checked="" type="checkbox"/>	Eric Trump says \$454m fine imposed on his father 'doesn't exist in this country'
<input checked="" type="checkbox"/>	RFK Jr. Apologizes for Super Bowl Ad ... That's Still Pinned Atop His Twitter Page
<input checked="" type="checkbox"/>	It's a date: Trump's first felony trial will be April 15
<input checked="" type="checkbox"/>	Joe Biden Pledges \$1.7 Billion To End Hunger Across U.S.

Sample Use Case: Misinformation Checker



r/politics

8.5m members ● 5.9k online



Select	Title
<input checked="" type="checkbox"/>	Biden Gets Some Good News in Poll as He Gains Ground on Trump
<input checked="" type="checkbox"/>	Biden calls for 'immediate ceasefire' in Gaza
<input checked="" type="checkbox"/>	Eric Trump says \$454m fine imposed on his father 'doesn't exist in this country'
<input checked="" type="checkbox"/>	RFK Jr. Apologizes for Super Bowl Ad ... That's Still Pinned Atop His Twitter Page
<input checked="" type="checkbox"/>	It's a date: Trump's first felony trial will be April 15
<input checked="" type="checkbox"/>	Joe Biden Pledges \$1.7 Billion To End Hunger Across U.S.

GENERATE
MISINFORMATION SCORE



Sample Use Case: Misinformation Checker



r/politics

8.5m members ● 5.9k online



Select	Title
<input checked="" type="checkbox"/>	Biden Gets Some Good News in Poll as He Gains Ground on Trump
<input checked="" type="checkbox"/>	Biden calls for 'immediate ceasefire' in Gaza
<input checked="" type="checkbox"/>	Eric Trump says \$454m fine imposed on his father 'doesn't exist in this country'
<input checked="" type="checkbox"/>	RFK Jr. Apologizes for Super Bowl Ad ... That's Still Pinned Atop His Twitter Page
<input checked="" type="checkbox"/>	It's a date: Trump's first felony trial will be April 15
<input checked="" type="checkbox"/>	Joe Biden Pledges \$1.7 Billion To End Hunger Across U.S.

GENERATE
MISINFORMATION SCORE



Sample Use Case: Misinformation Checker



r/politics

8.5m members ● 5.9k online



Select	Title	Misinformation %
<input checked="" type="checkbox"/>	Biden Gets Some Good News in Poll as He Gains Ground on Trump	40.83
<input checked="" type="checkbox"/>	Biden calls for 'immediate ceasefire' in Gaza	72.44
<input checked="" type="checkbox"/>	Eric Trump says \$454m fine imposed on his father 'doesn't exist in this country	57.28
<input checked="" type="checkbox"/>	RFK Jr. Apologizes for Super Bowl Ad ... That's Still Pinned Atop His Twitter Page	65.48
<input checked="" type="checkbox"/>	It's a date: Trump's first felony trial will be April 15	39.19
<input checked="" type="checkbox"/>	Joe Biden Pledges \$1.7 Billion To End Hunger Across U.S.	48.5

Sample Use Case: Misinformation Checker



r/politics

8.5m members ● 5.9k online



Select	Title	Misinformation %	Misinformation Alert
<input checked="" type="checkbox"/>	Biden Gets Some Good News in Poll as He Gains Ground on Trump	40.83	NIL
<input checked="" type="checkbox"/>	Biden calls for 'immediate ceasefire' in Gaza	72.44	ALERT
<input checked="" type="checkbox"/>	Eric Trump says \$454m fine imposed on his father 'doesn't exist in this country	57.28	ALERT
<input checked="" type="checkbox"/>	RFK Jr. Apologizes for Super Bowl Ad ... That's Still Pinned Atop His Twitter Page	65.48	ALERT
<input checked="" type="checkbox"/>	It's a date: Trump's first felony trial will be April 15	39.19	NIL
<input checked="" type="checkbox"/>	Joe Biden Pledges \$1.7 Billion To End Hunger Across U.S.	48.5	WARNING

Sample Use Case: Misinformation Checker



r/politics

8.5m members ● 5.9k online



Select	Title	Misinformation %	Misinformation Alert
<input checked="" type="checkbox"/>	Biden Gets Some Good News in Poll as He Gains Ground on Trump	40.83	NIL
<input checked="" type="checkbox"/>	Biden calls for 'immediate ceasefire' in Gaza	72.44	ALERT
<input checked="" type="checkbox"/>	Eric Trump says \$454m fine imposed on his father 'doesn't exist in this country'	57.28	ALERT
<input checked="" type="checkbox"/>	RFK Jr. Apologizes for Super Bowl Ad ... That's Still Pinned Atop His Twitter Page	65.48	ALERT
<input checked="" type="checkbox"/>	It's a date: Trump's first felony trial will be April 15	39.19	NIL
<input checked="" type="checkbox"/>	Joe Biden Pledges \$1.7 Billion To End Hunger Across U.S.	48.5	WARNING

Redditors
found
misleading

06

Conclusion & Recommendations

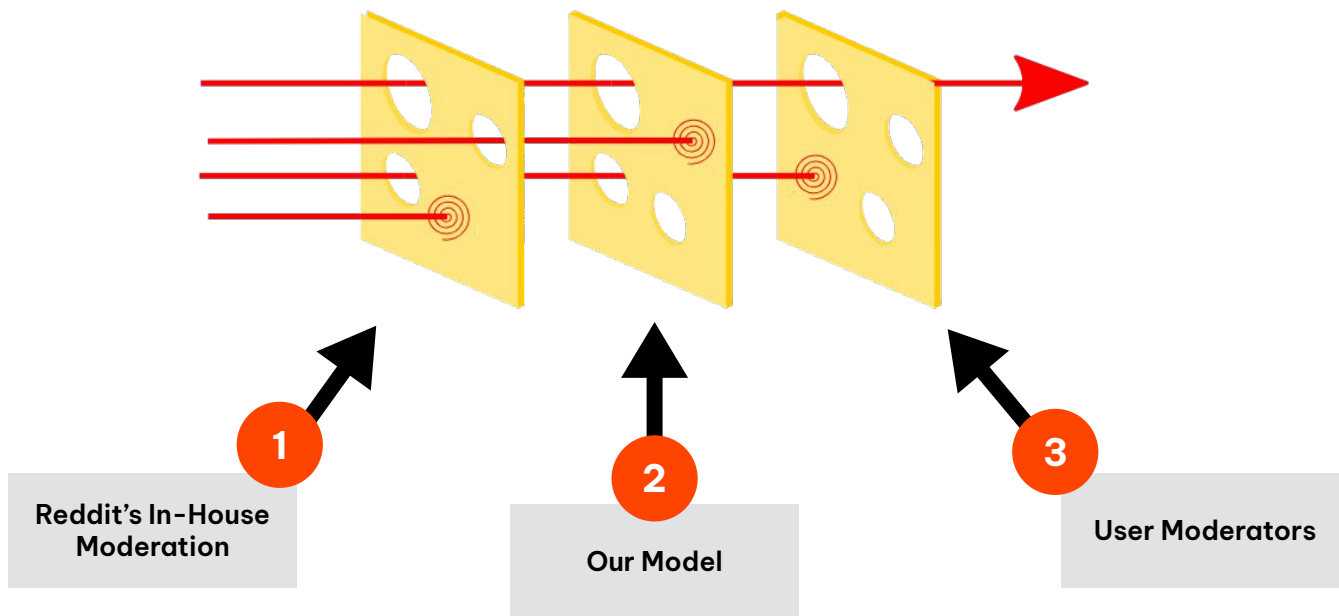
Conclusion

For a company that has gone newly public, Reddit has an **ever-increasing stake at content moderation** – and more specifically, **managing misinformation** – on its site.

In terms of scale, Reddit's in-house moderation is **easily dwarfed** by the **hundreds and thousands of posts** and **millions of active users** that the site sees daily. **Relying on its volunteer moderators** to create and enforce their own rules within each subreddit also **has its limitations**.

Recommendation

“The Swiss Cheese Model”



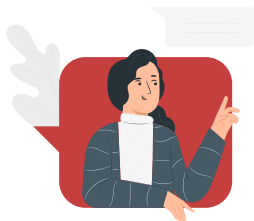
Potential Impact



Reddit

Primary Stakeholder

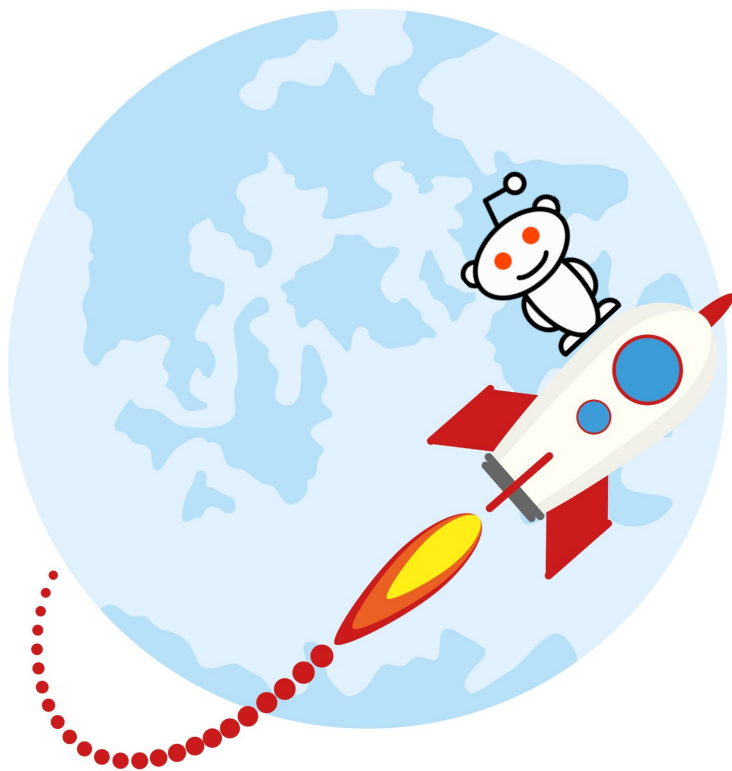
- Maintain its **brand status** as a reputable social media company
- “Swiss Cheese Model” could **work in tandem** with their existing content moderation strategy



User Moderators

Secondary Stakeholder

- Posts flagged as “misinformation” by our model would be **directed to the user moderators for review**
- User moderators may adopt a more **targeted approach** for content moderation that is **less resource-intensive**



Thank you

Q&A