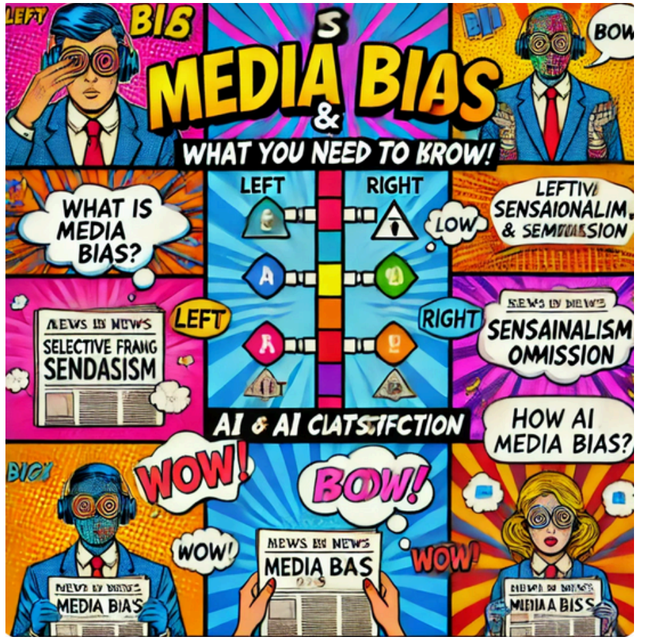
MACS 30100

Names

Research Question

TO WHAT EXTENT DO MEDIA SOURCES DIFFER IN THEIR FRAMING OF NEWS, AND HOW CAN MACHINE LEARNING BE USED TO CATEGORIZE STATEMENTS BY THEIR IDEOLOGICAL ALIGNMENT?

Media Bias ⟶ **Public Perception** ⟶ **Quality of democracy**



Literature Review

Fiore, V. (2019). Automatic identification of political ideology in online news articles, demonstrating automated classifiers can outperform human annotators in ideological categorization.

Chao, Z., Molitor, D., Needell, D., & Porter, M. A. (2022). Inference of media bias and content quality using natural-language processing with bidirectional LSTM neural networks.

Martín-Forero, Á. G., López, A. M., & Segura-Bedmar, I. (2022). UC3MDeep at PoliticEs 2022: Exploring traditional machine learning algorithms for political ideology detection.

Machine Learning Task

Multi-Class Classification
Three Classes--Left, Neutral, Right

Models
Random Forest, Logistic Regression, LLaMA, enhanced with ensemble, boost, and stacking

Allsides News Data

Source

- (Baly et al., 2020)

Outcome

- Article-Level Political Bias
- Left, Center, Right

Features & Preprocessing

- News articles crawled
- TF-IDF
- Text Embeddings

	topic	source	bias	title	date	authors	content	content_original	bias_text	ID	split	cloud_text	clean_text
0	coronavirus	The Guardian	0	US coronavirus stimulus checks: are you eligib...	2020-03-26	Lauren Aratani	Most taxpayers will get a check from the $ 2tn...	Most taxpayers will get a check from the $2tn ...	left	IBgtqEHUEcjyMriR	train	taxpayers get check $2tn economic package, lar...	Most taxpayers will get a check from the $ 2tn...
1	healthcare	Washington Times	2	After Obamacare health exchange deadline passe...	2013-02-16	Tom Howell Jr.	The backbone of President Obama ' s health car...	The backbone of President Obama's health care ...	right	LC3zdsbACLILzBhY	train	backbone President Obama's health care law tak...	The backbone of President Obama ' s health car...
2	politics	National Review	2	The DOJ Will Not Prosecute James Comey over Tr...	2019-08-03	Andrew C. Mccarthy, John Mccormack, Michael Br...	Then-FBI Director James Comey testifying on Ca...	Then-FBI Director James Comey testifying on Ca...	right	W084H19Vniu1Z24f	train	Then-FBI Director James Comey testifying Capit...	Then-FBI Director James Comey testifying on Ca...
3	media_bias	NPR Online News	1	Fox News Hasn't Always Shared Robert Driscoll'...	2018-08-27	David Folkenflik	Fox News Has n't Always Shared Robert Driscoll...	Fox News Hasn't Always Shared Robert Driscoll'...	center	Lcvv69hlu4J5Foc3	train	Fox News Always Shared Robert Driscoll's Crede...	Fox News Has n't Always Shared Robert Driscoll...
4	trade	Bloomberg	1	Wilbur Ross Says U.S., China 'Miles and Miles'...	2019-01-24	Brendan Murray, Andrew Mayeda	LISTEN TO ARTICLE 2:15 SHARE THIS ARTICLE Shar...	LISTEN TO ARTICLE 2:15 SHARE THIS ARTICLE Shar...	center	ZW7MmqXMtiJgL6jY	train	LISTEN ARTICLE 2:15 SHARE ARTICLE Share Tweet ...	LISTEN TO ARTICLE 2:15 SHARE THIS ARTICLE Shar...

df.shape
(36274, 15)

able	abortion	abortions	abroad	absolutely	abuse	accept	accepted	...	years
0.0	0.0	0.0	0.0	0.0	0.000000	0.0	0.000000	...	0.000000
0.0	0.0	0.0	0.0	0.0	0.000000	0.0	0.000000	...	0.042681
0.0	0.0	0.0	0.0	0.0	0.021804	0.0	0.000000	...	0.009380
0.0	0.0	0.0	0.0	0.0	0.000000	0.0	0.045837	...	0.000000
0.0	0.0	0.0	0.0	0.0	0.000000	0.0	0.000000	...	0.000000

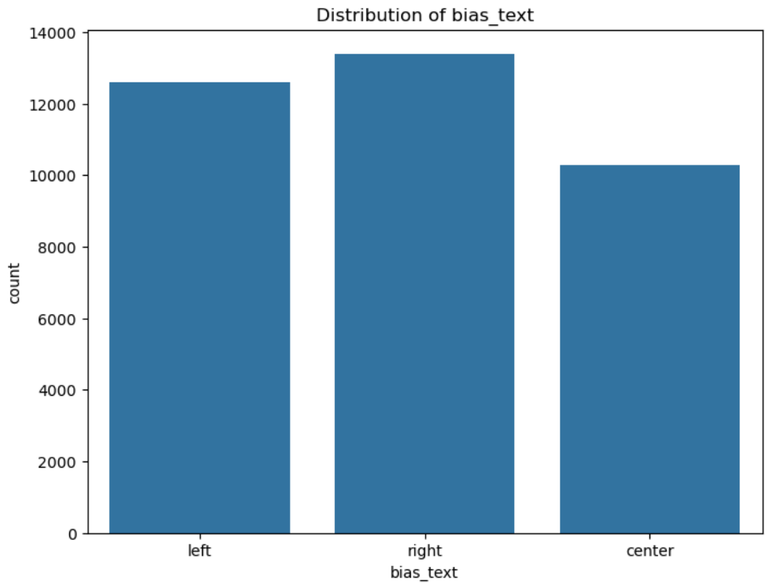
tfidf.shape
(36274, 3000)

	0	1	2	3	4	5	6	7	8	9	...
0	0.044637	0.010822	-0.031958	-0.023198	0.035853	-0.009816	0.020184	0.098970	0.022755	0.005069	...
1	0.009138	0.020595	-0.046163	-0.017980	0.030119	-0.008256	-0.015530	0.077065	0.026222	-0.034615	...
2	0.041385	0.011970	-0.035994	-0.012322	0.022936	-0.024032	-0.036985	0.074083	0.064832	-0.009455	...
3	0.025437	0.005591	-0.035518	-0.025949	0.031492	-0.024037	-0.015383	0.076002	0.034206	-0.003800	...
4	0.020295	0.021048	-0.039424	-0.001009	0.037378	-0.012874	-0.017437	0.085674	0.029621	-0.001029	...

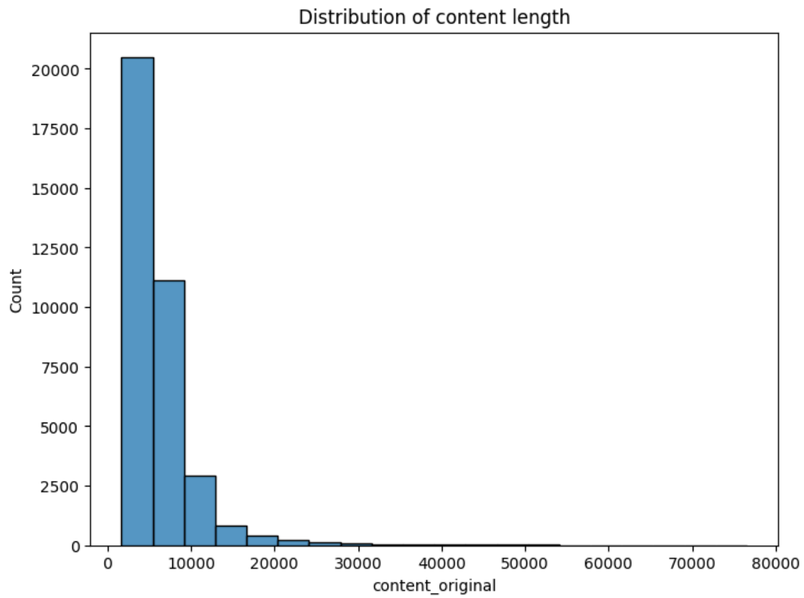
embedded.shape
(36274, 1024)

EXPLORATORY DATA ANALYSIS

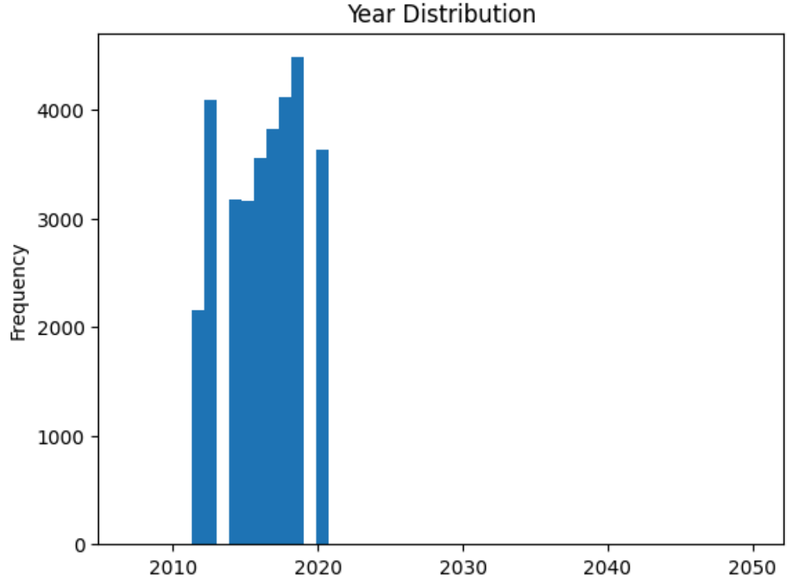
Distribution of Article Bias



Distribution of Article Length

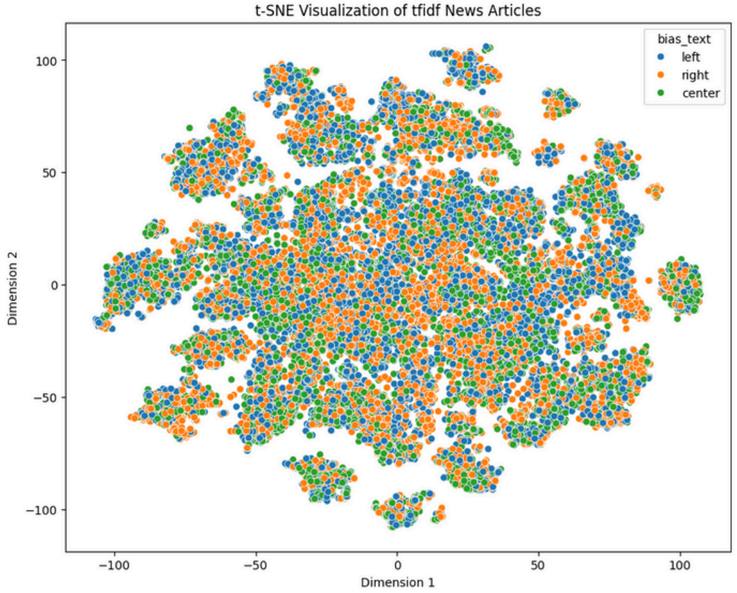


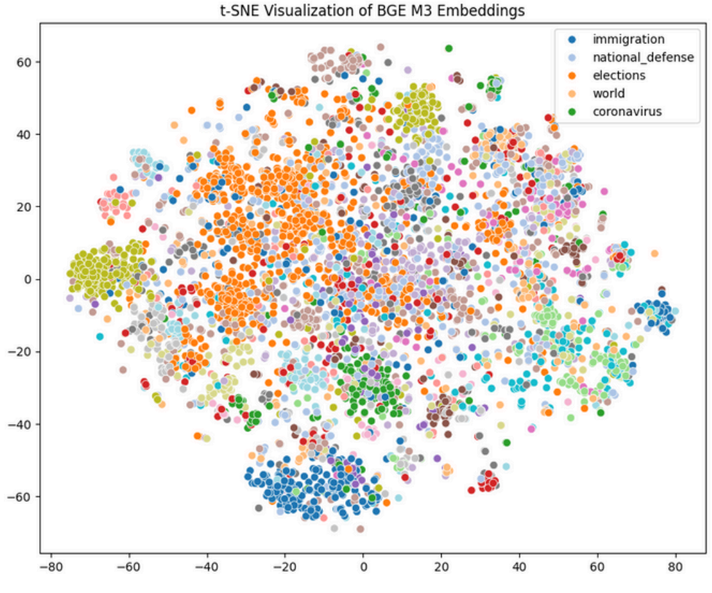
Distribution of Year

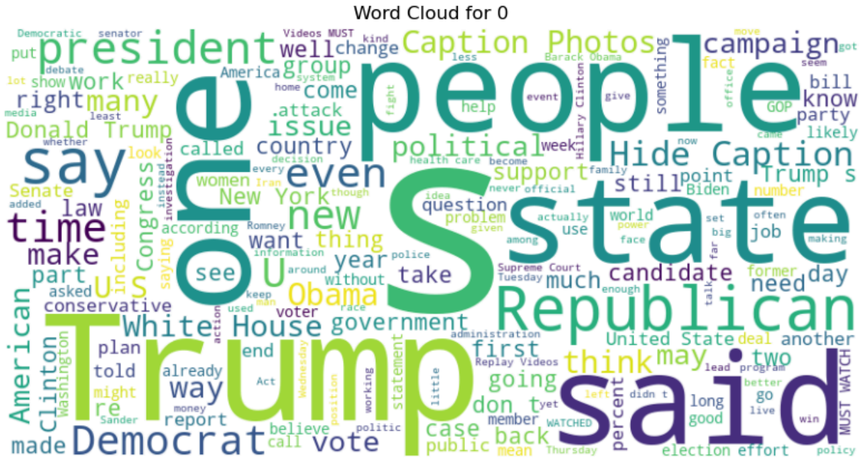


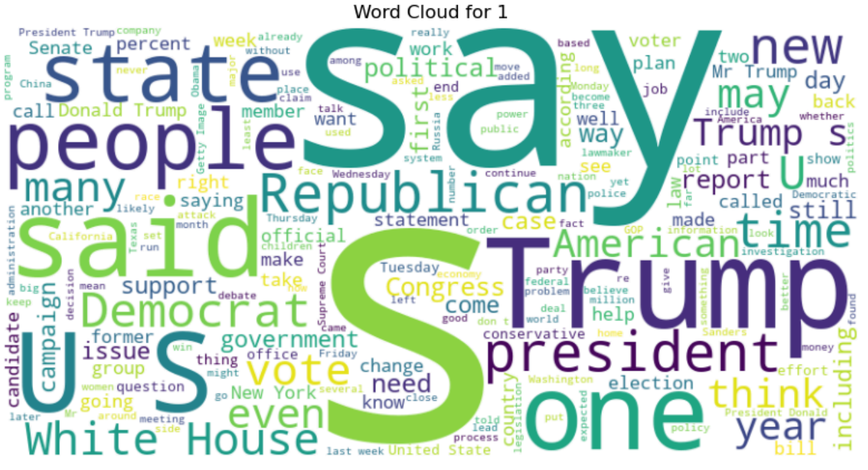
EXPLORATORY DATA ANALYSIS

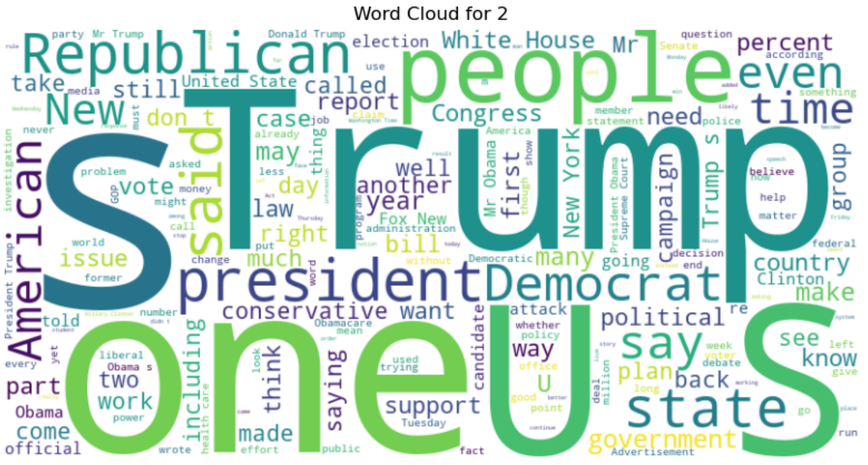
Left		Center		Right	
trump	0.068833	trump	0.090170	trump	0.064248
said	0.059145	said	0.066865	said	0.058229
obama	0.037310	president	0.040208	mr	0.049287
president	0.035066	mr	0.032708	president	0.040600
clinton	0.029868	says	0.032327	obama	0.037367
people	0.029843	house	0.032123	house	0.028516
house	0.029615	people	0.026684	clinton	0.026163
mr	0.028256	democrats	0.024989	people	0.026068
campaign	0.025520	new	0.024437	percent	0.025444
new	0.025287	state	0.024322	state	0.023815











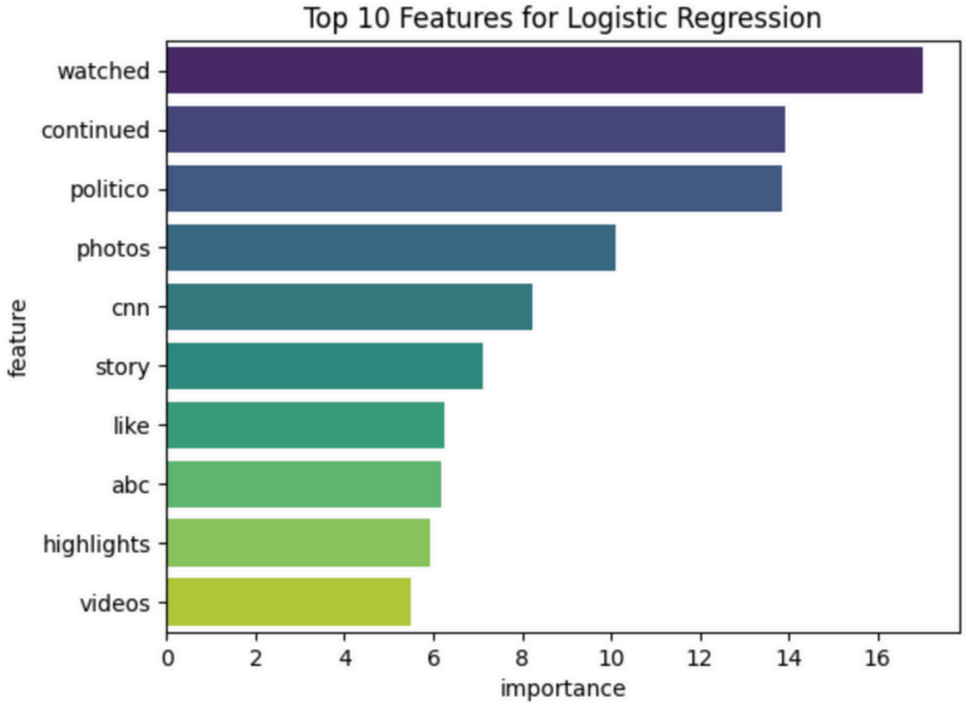
Logistic Regression

 # hyper-parameter tunning
param_rand = {
 'C': [0.1, 0.5, 1, 10, 100],
 'penalty':('l1', 'l2')}

RandomizedSearchCV(LogisticRegression(solver="saga", max_iter=1000, n_jobs=-1), param_rand,cv=2, verbose=2, n_iter=10, n_jobs=-1)

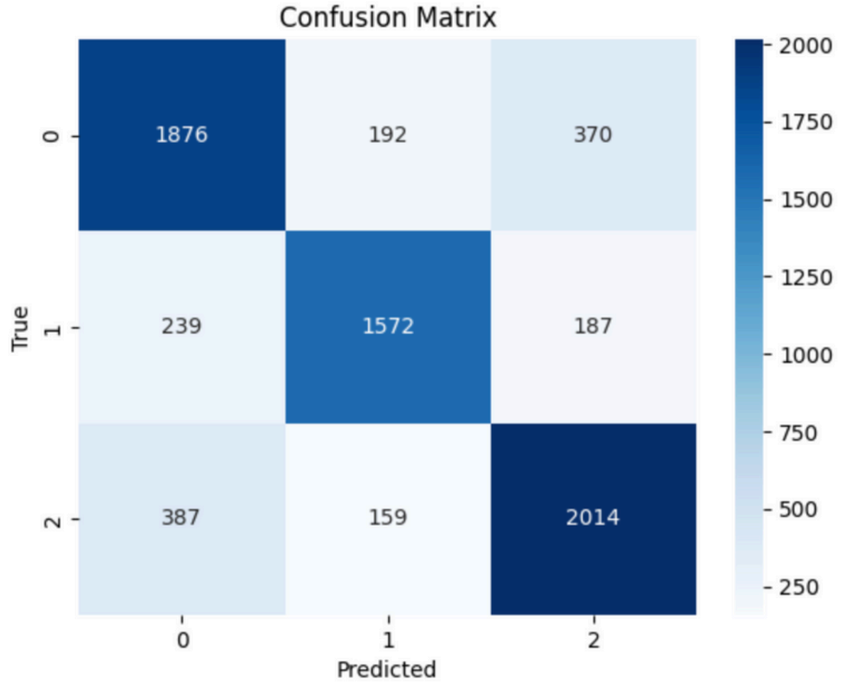
 # best hyperparameters
{'penalty': 'l1', 'C': 1}

 # no. nonzero coefficients
2185



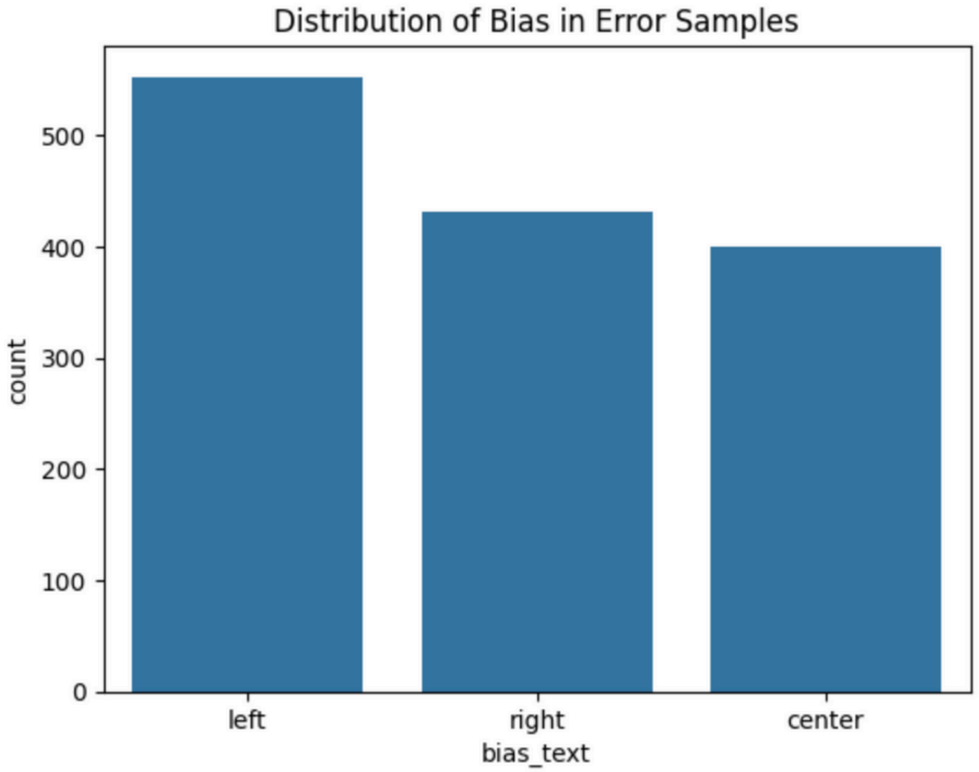
After tuning

	precision	recall	f1-score	support
left	0.75	0.77	0.76	2438
center	0.82	0.79	0.80	1998
right	0.78	0.79	0.79	2560
accuracy			0.78	6996
macro avg	0.78	0.78	0.78	6996
weighted avg	0.78	0.78	0.78	6996



Logistic Regression

& Error Analysis



Selected wrong prediction

> The Trump administration and its Republican enablers are fighting a series of wars directed at **targets inside the United States**...US government officials, especially but not exclusively Republicans, have been **lying to the American people** about matters of life and death for a long time....Anyone reading with a critical eye would have recognized his argument's **weakness** and dismissed it as **deranged fascist propaganda**.

- the classifier predicts the bias of the piece as **right**
- yet, the article presents a strong anti-Trump stance, using highly charged language and framing his administration as an threat to democracy

Random Forest

Before Tuning

	precision	recall	f1-score	support
left	0.74	0.78	0.76	2438
center	0.86	0.77	0.81	1998
right	0.76	0.79	0.78	2560
accuracy			0.78	6996
macro avg	0.79	0.78	0.78	6996
weighted avg	0.78	0.78	0.78	6996

After Tuning

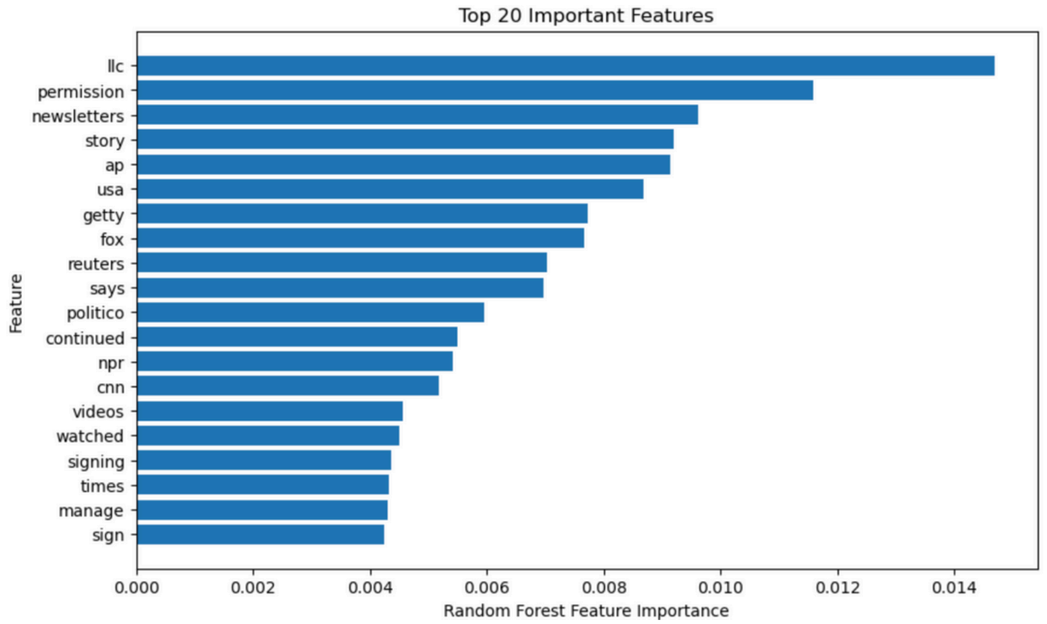
	precision	recall	f1-score	support
left	0.79	0.78	0.79	2438
center	0.88	0.80	0.83	1998
right	0.78	0.84	0.81	2560
accuracy			0.81	6996
macro avg	0.81	0.81	0.81	6996
weighted avg	0.81	0.81	0.81	6996

Tuning by randomized search

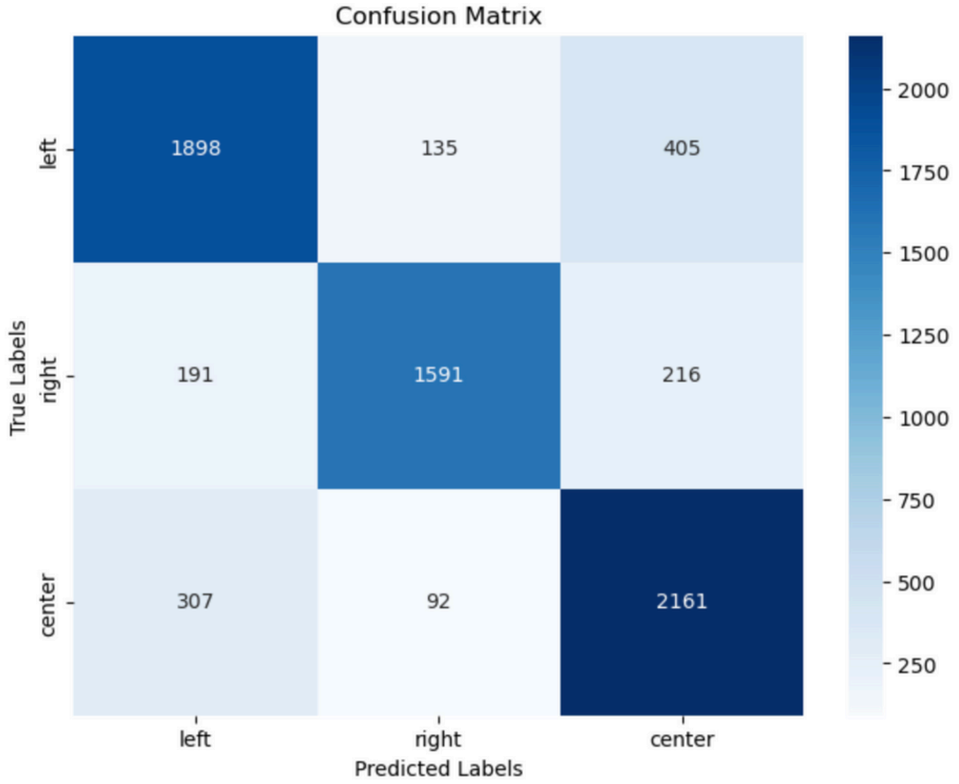
```
param_grid = {
    'n_estimators': [100, 300, 500, 1000],  # Number of trees
    'max_depth': [30, 50, 100, 125, 150],     # Depth of trees
    'min_samples_split': [2, 5, 10]   # Minimum samples per split}
random_search=RandomizedSearchCV(RandomForestClassifier(ran
dom_state=42, n_jobs=-1), param_grid, cv=3, verbose=2, n_iter=15)
```

Best hyperparameters

Best Parameters: {'n_estimators': 1000,
'min_samples_split': 2, 'max_depth': 125}
Tuned Random Forest Accuracy:
0.8037449971412236



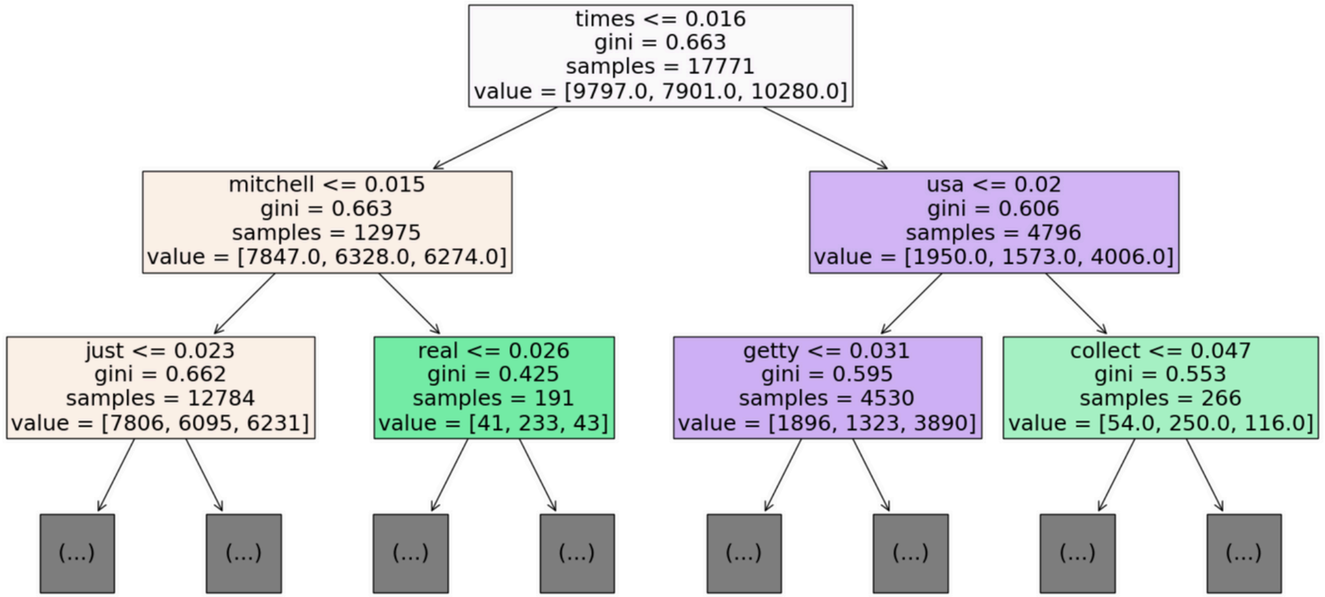
Random Forest



Pro & Con

High accuracy
High recall for right
Balanced precision
Resistant to Overfitting

Low precision for “Right”
Fail to classify highly overlapping feature



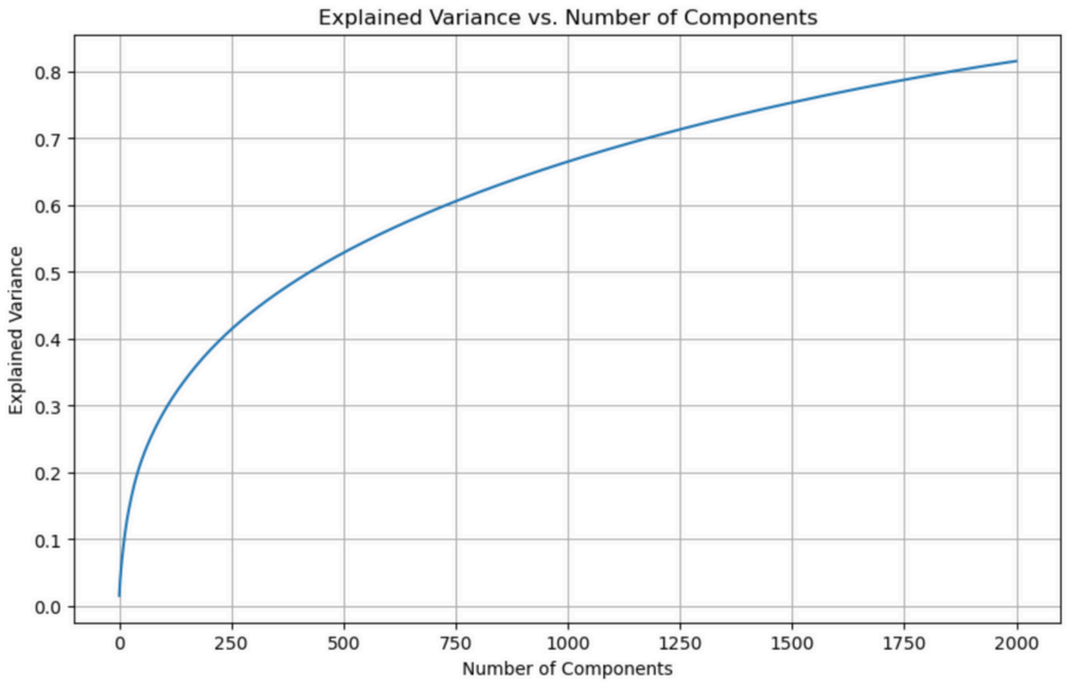
Random Forest

After PCA

	precision	recall	f1-score	support
left	0.59	0.55	0.57	2438
center	0.76	0.29	0.42	1998
right	0.53	0.81	0.64	2560
accuracy			0.57	6996
macro avg	0.63	0.55	0.54	6996
weighted avg	0.62	0.57	0.55	6996

PCA

A linear dimensionality reduction technique
Text data is represented as high-dimensional and sparse
Loss of Important Features & Word Relationships
Help with computation speed



LlaMA + QLoRA

Hyperparameters

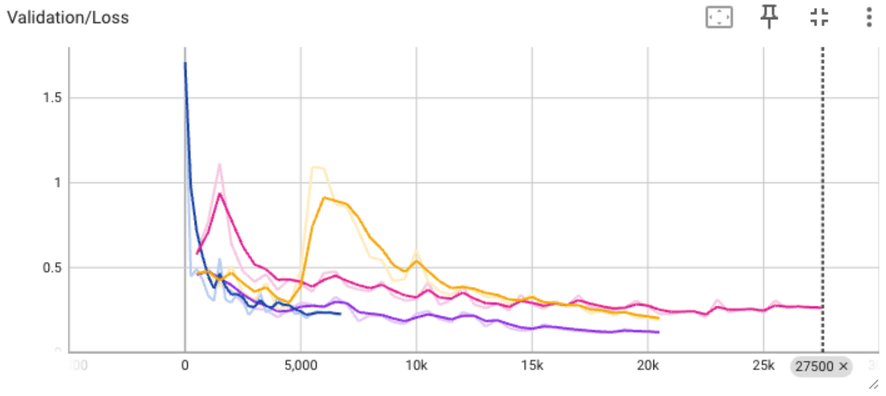
- **LoRA Rank:** 8, **16**
- **LoRA Alpha:** 16, **32**
- **LoRA Dropout:** 0.05, **0.1**
- **Batch Size:** 2, **4**, 16
- **Epoch:** 3, **4**

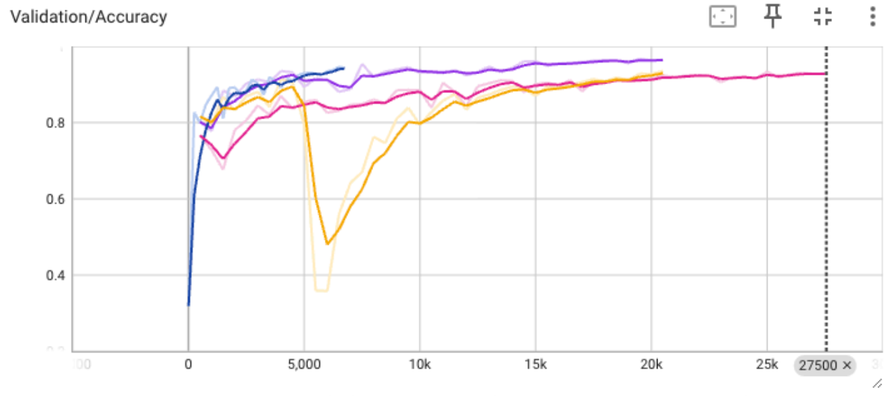
Training

- Larger LoRA Rank needs larger batch size
- Larger Batch size needs more steps to converge
- A balance is achieved at r16+b4

	precision	recall	f1-score	support
left	0.90	0.95	0.92	402
center	0.96	0.98	0.97	299
right	0.97	0.92	0.95	599
accuracy			0.94	1300
macro avg	0.94	0.95	0.95	1300
weighted avg	0.95	0.94	0.94	1300

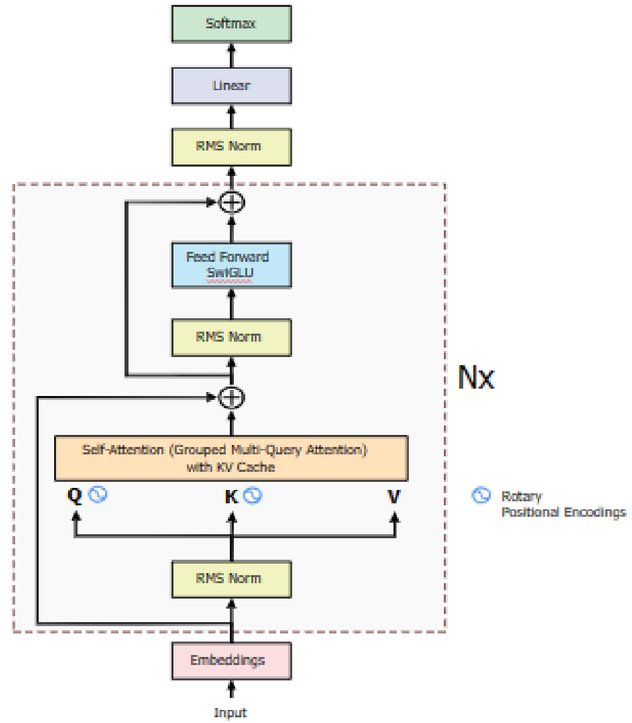
Classification Report (Test)

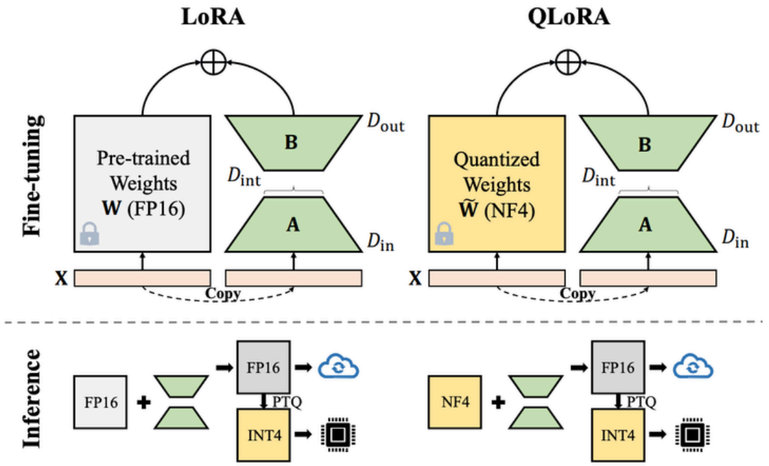




Run	Smoothed	Value	Step	Relative
exp-r16-b16-e4	0.9417	0.9434	6,750	4.745 hr
exp-r16-b2-e3	0.9295	0.9355	20,500	6.679 hr
exp-r16-b4-e4	0.964	0.9648	20,500	2.977 hr
exp-r8-b2-e4	0.9281	0.9287	27,500	8.818 hr

LlaMA 3





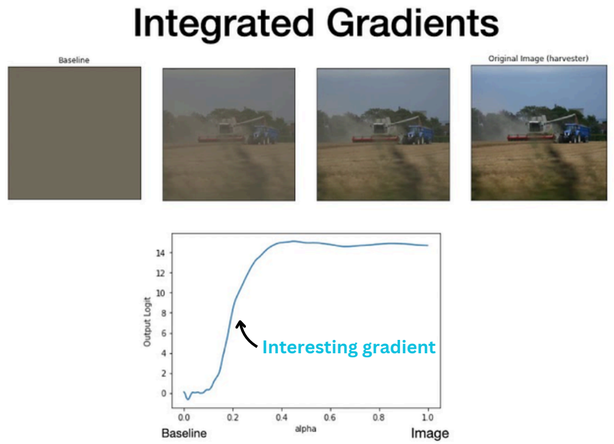
LlaMA Captum Analysis

Legend: Negative Neutral Positive — Word Importance

During an April 2 interview with Chris Wallace on Fox News, Surgeon General Jerome Adams compared deaths caused by COVID-19 to deaths caused by smoking and drug abuse. "More people will die, even in the worst projections, from cigarette smoking in this country than are going to die from coronavirus this year," he said. Wallace questioned the analogy. "Dr. Adams," he said, "there's a big difference between opioids and cigarettes, which are something that people decide to use or not to use, [and] the coronavirus, which people catch. It's not an individual choice." The distinction that Wallace considered commonsensical is not one that public health officials like Adams recognize. As they see it, their mission is minimizing "morbidity and mortality," whether those things are caused by communicable diseases or by lifestyle choices. Equating true epidemics with metaphorical "epidemics" of risky behavior distracts public health agencies from their central mission of protecting people against external threats such as pollution and pathogens. It undermines their moral authority by implying that the rationale for that uncontroversial mission also justifies a wide-ranging paternalism, and it damages their credibility by involving them in high-handed, manipulative propaganda. The ambiguity about what it means for the government to protect public health also makes it harder to think clearly about the limits of state power in responding to literal epidemics. The classical liberal tradition has always recognized that the state has a legitimate role to play in protecting the public from contagious diseases. When we are confronted by an actual public health crisis like the COVID-19 pandemic, the question is not whether the use of force can be justified but whether a particular policy is appropriate. That question is hard to answer when there is a high degree of uncertainty about the threat posed by the disease and the cost of limiting its spread. The Vaping 'Epidemic' Before the new coronavirus came along, the U.S. Centers for Disease Control and Prevention (CDC) spent a lot of time and effort warning us about a very different kind of "epidemic": an increase in e-cigarette use by teenagers, coupled with an outbreak of vaping-related lung injuries. The first concern did not involve any sort of disease; the latter did, but unlike COVID-19, the condition that the CDC dubbed "e-cigarette, or vaping, product use-associated lung injury" (EVALI) was not contagious. And contrary to the CDC's misleading nomenclature and dangerously misguided initial advice, the two developments appeared to be completely unrelated. Even under a broad understanding of public health, the CDC's conflation of EVALI with vaping in general was counterproductive, impeding the harm-reducing shift from conventional cigarettes to nicotine delivery systems that are far less dangerous. By fostering confusion about the relative hazards of smoking and vaping, the CDC hurt its own credibility on the eve of a public health crisis in which policy makers and the rest of us were expected to rely on its expertise. As Chris Wallace probably would agree, vaping is something people choose to do. In that respect it resembles many other phenomena that politicians, bureaucrats, and academics have described as public health problems, including smoking, drinking, illegal drug use, overeating, physical inactivity, riding a motorcycle or bicycle without a helmet, gambling, playing violent video games, and watching pornography. COVID-19, by contrast, is something

A rightist article criticizing CDC misclassified as left-leaning

Green Words such as Jerome and COVID-19 pushed the prediction to the left



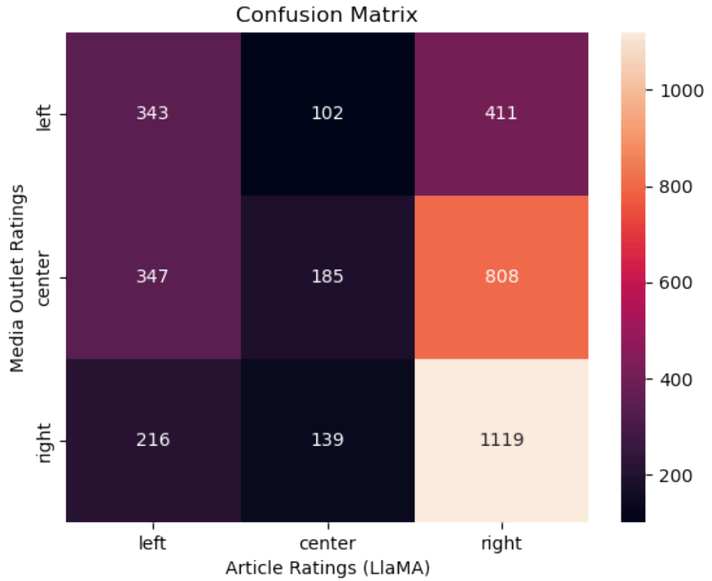
$$\text{Interpolated Embedding}(k) = \text{Baseline} + \frac{k}{N} \times (\text{Input} - \text{Baseline})$$

Captum Integrated Gradients iteratively moves each input token embedding from the baseline token (e.g., padding token) to the original embedding, accumulate the gradient through backpropagation and decide the importance of the token

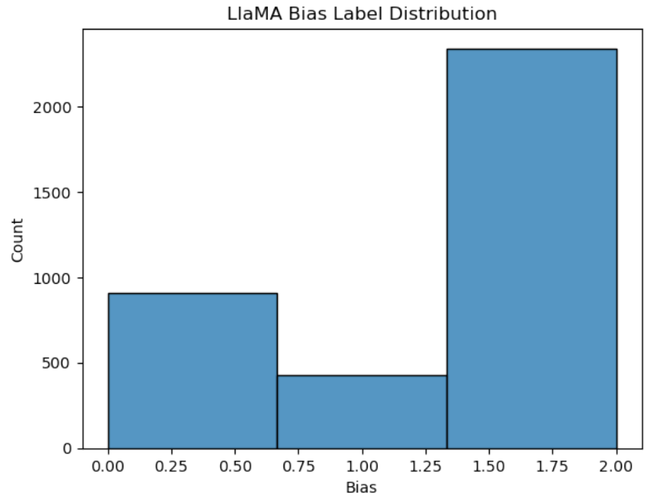
LlaMA + QLoRA

Classifying news in the Wild

- 5770 news stories
- 1450k news articles from Ground News
- 10k samples



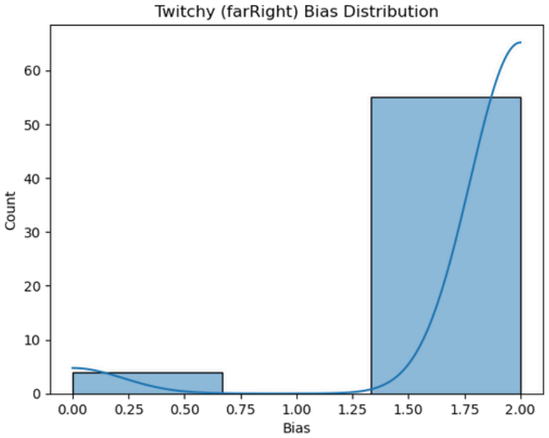
Outlet Ratings vs. LlaMA Ratings (Ground News 10k)

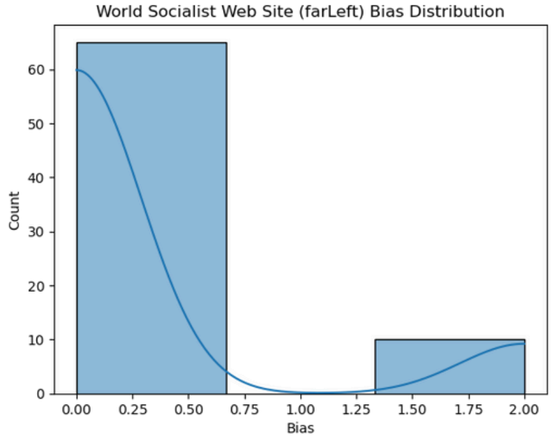


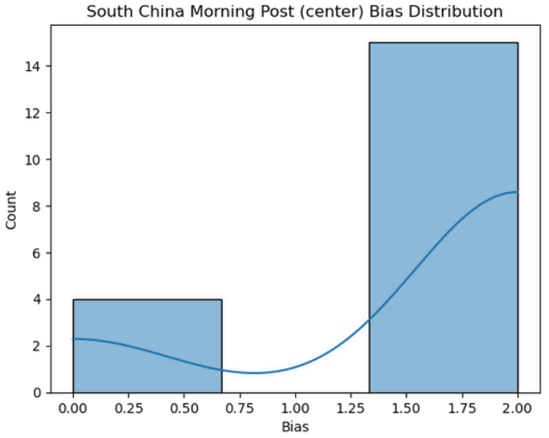
LlaMA Rating Distribution

Analysis

- The prediction is skewed to the right
- LlaMA ratings align well with some outlet ratings, but not all







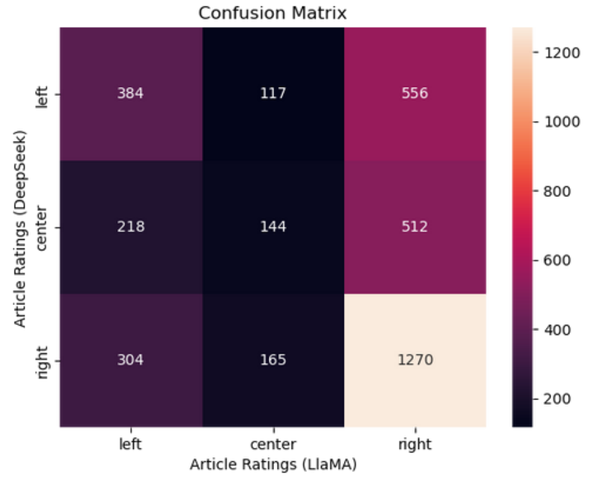
LlaMA vs. DeepSeek

```
**SYSTEM:**

It is {current_time}. You are an expert analyst in political media analysis.
Your task is to analyse how the author or article's view align with the
left, center or right-wing policies or perspectives in the United States
around 2020s, and provide the labels for the article according to its
alignment.

Follow the instructions below:

1. **Reasoning**:
    1. **Potentially Biased Coverage**: List any potential partial coverage,
    biased quotations, or author's subjective opinion in bullet points.
    2. **Alignment with US Political Ideologies**: Explain how these align
    with left, center or right-wing policies or perspectives in the US around
    2020s.
    3. **Overall Assessment**: Aggregate the information to provide an
    overall assessment of the article's political bias.
2. **Political Compass Score**: A numerical float value between -2.0 (left)
and 2.0 (right).
3. **Categorical Label**: Choose one from (`left`, `left-center`, `center`,
`center-right`, `right`).
```

Prompt for DeekSeek R1

Comparison

- LlaMA **skewed to the right across the board** when compared with DeepSeek labels
- Two models has a cohen's kappa score of **0.15, slight agreement**
- Calls for a better dataset for political bias detection



KAPPA CALCULATION

$$\kappa = \frac{Pr(a) - Pr(e)}{1 - Pr(e)}$$

Pr(e) Calculation

$$\text{Expected Agreement} = \frac{\left(\frac{cm^1 \times rm^1}{n}\right) + \left(\frac{cm^2 \times rm^2}{n}\right)}{n}$$

$$\text{Expected Agreement} = \frac{\left(\frac{157 \times 150}{222}\right) + \left(\frac{65 \times 72}{222}\right)}{222}$$

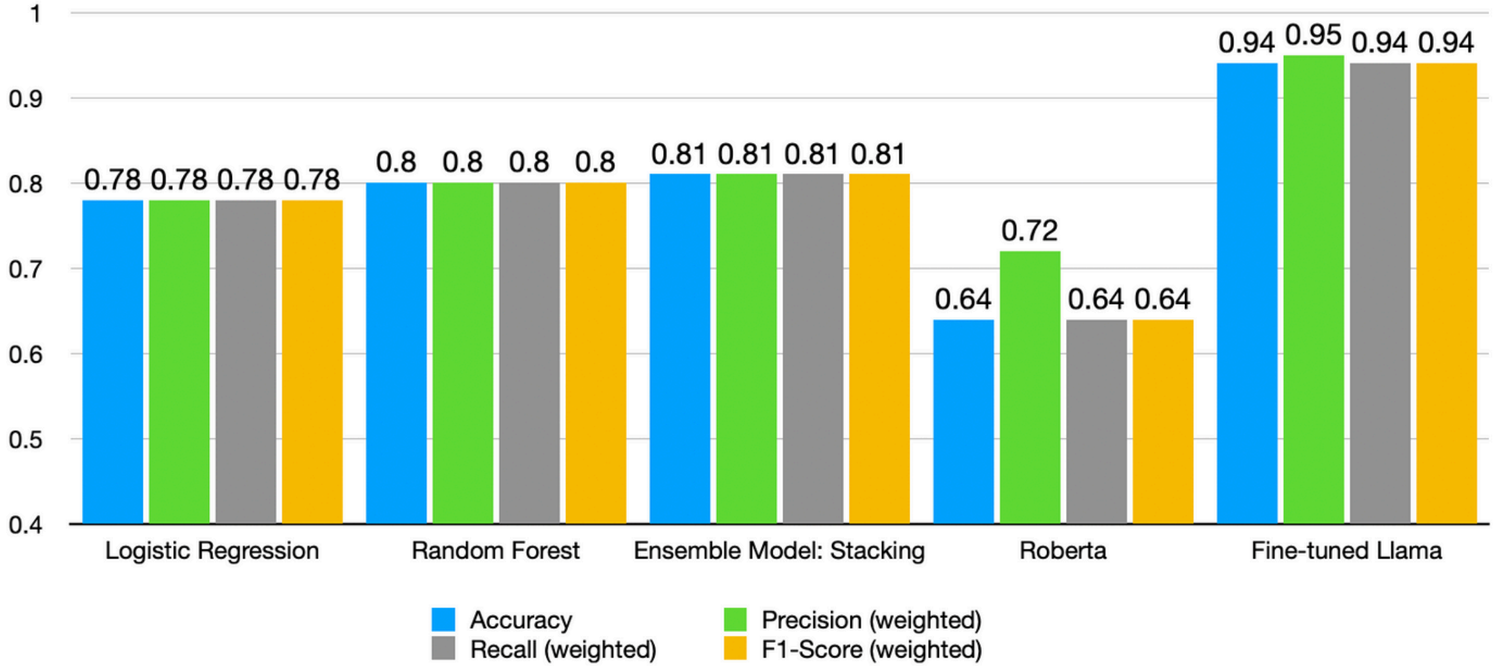
$$\text{Expected Agreement} = \frac{108.08 + 21.08}{222} = .57$$

$$\text{Kappa} = \frac{.94 - .57}{1 - .57} = .85$$

	precision	recall	f1-score	support
left	0.42	0.36	0.39	1057
center	0.34	0.16	0.22	874
right	0.54	0.73	0.62	1739
accuracy			0.49	3670
macro avg	0.44	0.42	0.41	3670
weighted avg	0.46	0.49	0.46	3670

Feingold, M. (1992). The equivalence of Cohen's Kappa and Pearson's chi-square statistics in the 2× 2 table. Educational and psychological measurement, 52(1), 57-61.

Model Comparison & Evaluation

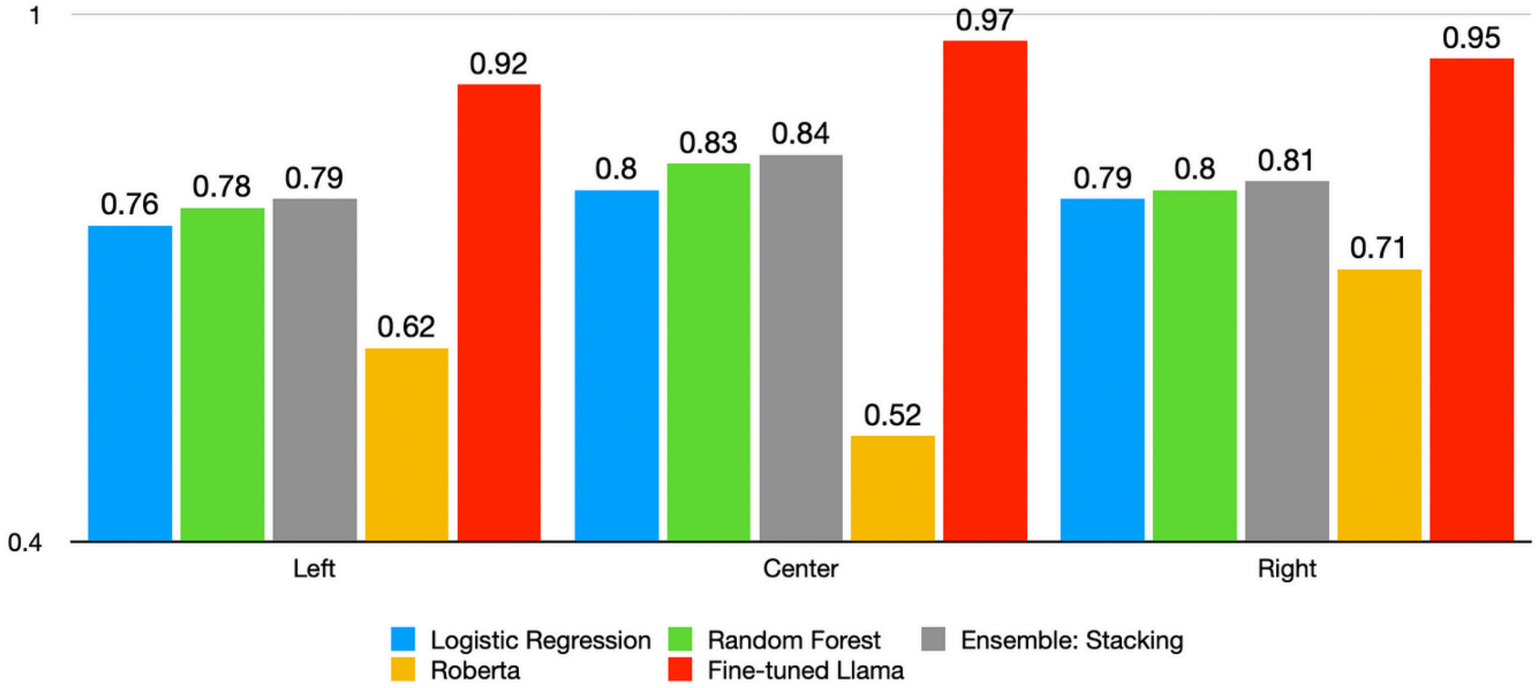


Weighted Average Performance

Model Performance Comparison

	Accuracy	Precision (weighted)	Recall (weighted)	F1-Score (weighted)
Logistic Regression	0.78	0.78	0.78	0.78
Random Forest	0.8	0.8	0.8	0.8
Ensemble Model: Stacking	0.81	0.81	0.81	0.81
Roberta	0.64	0.72	0.64	0.64
Fine-tuned Llama	0.94	0.95	0.94	0.94

Model Comparison & Evaluation



Per-class Performance Comparison

Per Class Prediction Performance

	Logistic Regression	Random Forest	Ensemble: Stacking	Roberta	Fine-tuned Llama
Left	0.76	0.78	0.79	0.62	0.92
Center	0.8	0.83	0.84	0.52	0.97
Right	0.79	0.8	0.81	0.71	0.95