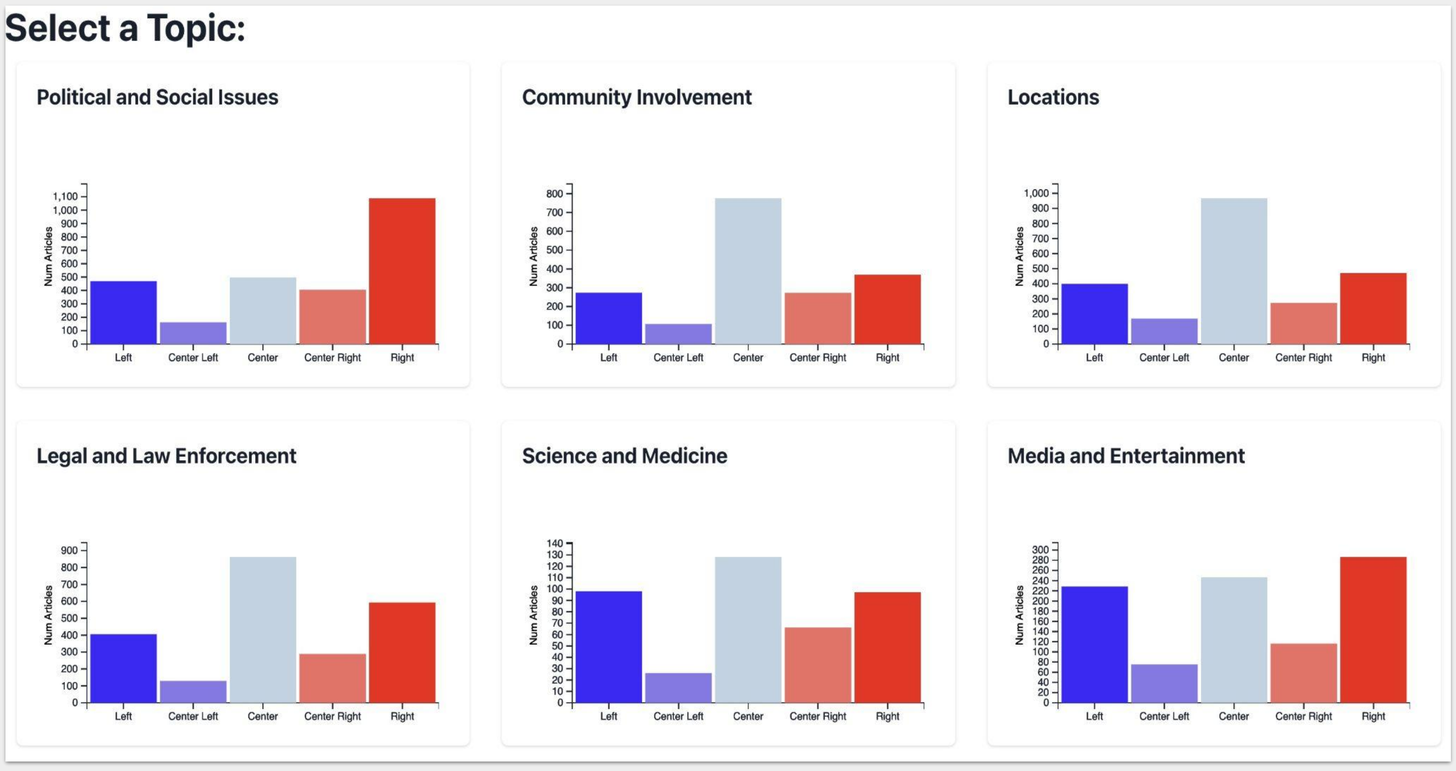
NewsMatch: Political Bias Recommendation and Visualization

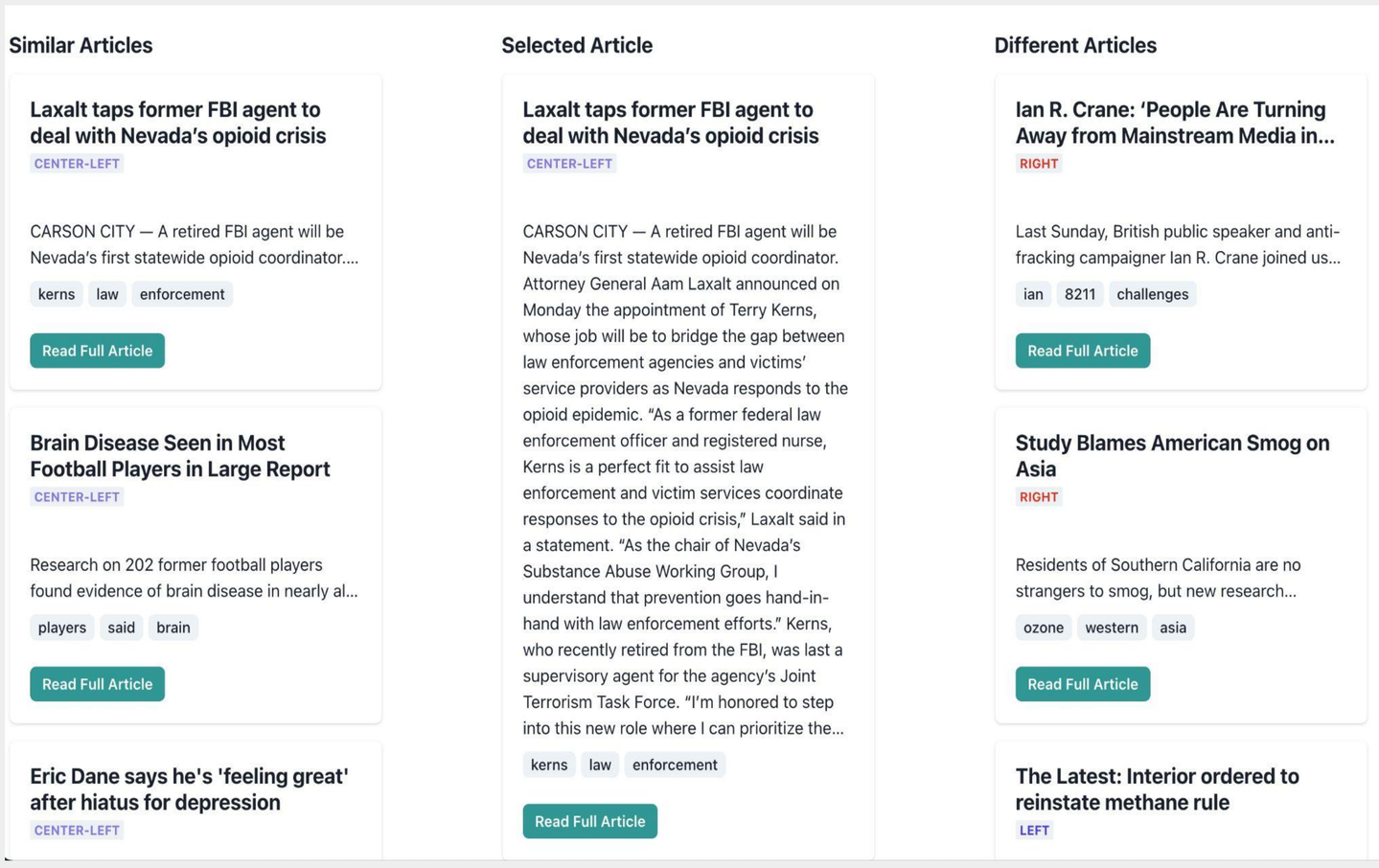
Andrew Zhang, Bryan Zhao, Neel Ganediwal, Yash Gupta, Aayush Seth, Shiva Ramaswami
Team #15

Introduction and Motivation

NewsMatch is an interactive tool that can **detect and visually convey bias** in news articles. The system also recommends users articles with a wide variety of viewpoints, promoting a **more holistic view of the news**. With the upcoming 2024 American presidential elections, it is important that voters are able to make informed decisions. Many people only look for information that agrees with their viewpoints without considering opposing viewpoints. This can also help tackle the growing problem of uninformed voters.



Our NewsMatch interface displays extracted topic clusters and the distribution of bias within those respective topics (Right). The system **identifies potential coverage bias**, clearly showing if a topic receives uneven coverage from certain perspectives. After selecting a topic, the user is presented with a list of headlines, where they can select an article to view. The system then displays bias labels and recommendations other articles with different biases within the same topic cluster (Bottom), promoting **diverse viewpoints and perspectives**.



Our Approach

Bias Classification: Use **DistilBERT** to encode articles into feature vectors and experiment with various classification models to classify **fine-grained bias labels**.

Keyword Extraction: Use the **PositionRank** algorithm to summarize each article into 5 keywords.

Keyword Clustering: Use **word2vec** to encode keywords and **KMeans** to group word vectors into broader **topic clusters** which are then and assigned to corresponding articles.

Recommendation System: Interactively present articles by topic and display distribution of biases to convey potential **coverage bias**. When the user selects an article, our system provides recommendations based on the predicted bias labels, showing both **similar and opposing viewpoints**.

Data

Use the **Hyperpartisan News Detection** dataset from HuggingFace containing over **750k news articles** with **fine-grained bias labels** ranging from 0-4, representing viewpoints ranging from extreme left to extreme right. Article text was preprocessed to remove HTML artifacts.

Experiments and Results

Bias Classification

Our best performing model consists of a fine-tuned DistilBERT with a fully-connected layer for sequence classification, achieving an **accuracy of 0.891** and **mean absolute error of 0.294**. We found that fine-tuning DistilBERT instead of using the frozen model to encode articles into feature vectors resulted in a significant increase in performance.

	Model	Accuracy	MAE
Frozen	SGDClassifier	0.411	1.457
	Perceptron	0.404	1.549
	MultinomialNB	0.375	1.360
	BernoulliNB	0.245	1.287
	PassiveAggressive	0.403	1.471
	FCN Head	**0.692**	**0.733**
Fine-Tuned	FCN Head	**0.891**	**0.294**

Keyword Clustering

DBScan was found to create uneven clusters, and was deemed unsuitable for our use case. The elbow method was used to determine the optimal number of 14 clusters for KMeans, of which 5 consisted of non-cohesive or general verbs unsuited for topic assignment. The remaining **9 clusters** were then manually assigned topic names, listed below.

- Political and Social Issues
- Community Involvement
- Locations
- Legal and Law Enforcement
- Science and Medicine
- Media and Entertainment
- Nature and Wildlife
- Food
- Finance and Economics

While our clustering achieves slightly subpar metrics with a **Silhouette Coefficient of -0.0697** and **DB-Index of 5.2527**, we find that the clustering functions well qualitatively and **provides broad topic clusters** that can be effectively used in our system to group and categorize articles.

User Studies

A small-scale user study performed among acquaintances asked participants to rate their agreement statements on a scale from 1 to 5. Based on the responses shown to the right, we conclude that our system is initially effective in both informing users of news bias and encouraging readers to explore opposing viewpoints.