News Bias Detector: A Platform to Detect Bias and Segregate News Content based on Keyword

Saurav Bhattacharyya | Aachal Thapa | Peerada Looareesuwan

Abstract – **With the wide variety of news content available on the Internet, finding authoritative sources of information is important to readers. The news can also be biased based on where the news source is from. The bias can also be viewed on social media. Readers are interested to read news from certain news sources or social media, with particular topics. Noteworthy, they may also have some personal biases while choosing the source of news. This paper introduces a framework that categorizes the news into certain bias buckets and lets consumers choose the type and topic of news that they want to read. Based on this preference the news would be served.**

Keywords – **News, Media influence, Bias.**

I. INTRODUCTION

The world can be viewed as a continuous stream of events, confusingly complex and entangled. From this huge stream a tiny part is selected, 'torn apart' by news correspondents and sent as 'news' to newspapers and broadcasting stations throughout the world [1]. Media influences our daily decisions. Our lives are constantly being driven by what we see and what we hear. Media is a source of information, and if people agree with what they hear, they are more likely to be subject to "conformation bias." In addition, in politics, news and media play an important role in influencing the people and their opinion, therefore influencing whom they may vote for in an election [2].

Sometimes news channels or sources can be biased with news that purportedly distorts or falsifies reality, sometimes to news that favors one side rather than providing equivalent treatment to both sides in a political conflict and sometimes to the motivations and mindsets of journalists who allegedly produce the biased content. Most of the studies that do explicitly explore bias focus on presidential campaigns and administrations and find little evidence of decisive or consistent, liberal or conservative, Democratic or Republican bias [3].

Media bias is the bias or perceived bias of journalists and news producers within the mass media in the selection of events and stories that are reported and how they are covered. Now that we know these biases exist, it is important to recognize whether or not bias gets in the way of truthful reporting. We should know how media influences our options and decisions to make sure we are aware of what media sites we pay attention to and what information we might be missing out on [4].

II. RELATED WORK

In a political climate where sharing fake news has become commonplace, it's more important than ever to rely on trustworthy and dogged fact-checking services to vet information [5]. There are some fact checker sites in the web and some google scholar article on media bias.

Fact checker is an application which screens the accurate precision of what is said by major U.S. political players as TV advertisements, discusses, discourses, meetings and news discharges.

Washington Post Fact Checker usually have left-center bias. They fact check conservative claims more than liberal ones.

ProPublica produces investigative journalism with moral force about important issues that occurs in the world.

This paper presents a system that classifies the news into certain predisposition containers as liberal and conservative category and gives users a chance to pick the sort and subject of news that they need to peruse which makes a major difference between these fact checker applications and our project.

III.RESEARCH QUESTIONS

1. How to categorize the news channel as right, left or middle? - We take help from several established research paper, conducted on news media and determine the bucket they belong to. We would have a weight corresponding to each news channel. For example, fox news is 100% conservative and that Huffington post if 100% liberal.
2. How to identify biases from different news channel? - Given any topic, the content is extracted from a news channel. The news initially marked as the same weight as the news channel itself. We also keep a master list of keywords, used by conservative vs liberal media, that would categorize the content bias.
3. What sources can be used for cross validation (AWS comprehend, twitter stream etc.)? - We would use amazon comprehend for keyword extraction and sentiment analysis purposes. We would also use internally collected keyword match.
4. How to cross validate the biases of a news? - The news links are extracted as part of RSS feed. The links represent the full content which will be extracted and send to amazon comprehend system. AWS comprehend would extract and give us back the keywords. The keywords would be internally matched with a master list of keywords to determine the bias of the content itself.
5. Where is the news content link stored and retrieved? - The links are extracted from the RSS feed. The project architecture would propose storing the links (but not the content) in AWS cloud.
6. Where and in what format the result is stored? - We store the data in database in a relational format.
7. How to make users aware about the news biases? - User will be presented with a preference page where they can choose their content selection they can also pick and choose the type of news they would like.

8. What can users do with the recommendations/biases? - User can go back and change their preferences and can get more/less result.

IV. PROJECT DESCRIPTION

Technical Description: The project consists of several sub systems each communicating with each other. the sub systems are:

a. Cloud DB:
   This is the central DB that holds the following
   - The preferred news channel and their relative biases.
   - The weightage of the news channel
   - The result of the news biases.
   - User preferences

b. RSS Feeder:
   The RSS feeder is the central engine that calculates bias. This pulls the RSS feed from the pre-configured news channel. The links are extracted, and the full story is extracted from the news source. The text is run against the amazon comprehend. The key phrases are extracted and compared against the master key list. Once compared, the news is calculated based on the number of matched key phrases. The matching keys have a bias and thus can be categorized either liberal or conservative. The calculation results are stored in the cloud DB.

c. Biased Source:

The Huffington Post	**L**	𝕃	ℂ	ℝ	ℝ
MSNBC	**L**	𝕃	ℂ	ℝ	ℝ
BBC News	𝕃	𝕃	**C**	ℝ	ℝ
ABC News	𝕃	**L**	ℂ	ℝ	ℝ
Fox News	𝕃	𝕃	ℂ	**R**	ℝ
BreitBart	𝕃	𝕃	ℂ	ℝ	**R**

d. Keywords Cloud:

	Liberal	Conservative
Abortion:	Fetus	Unborn baby
Affirmative Action:	Unequal opportunity	Discrimination
Climate Change:	Carbon dioxide	Natural

Corporate:	Corporate transparency	Corporate accountability
Customer law:	Consumer protections	Corporate regulations
Death Penalty :	Imprisonment	Executing a murderer
	Cruel and unusual	Crime of murder
	Risks killing	Punishment
Economy:	Global economy	Free market economy
Energy:	Drilling for oil	Exploring the energy
Euthanasia	Right to die	Immoral
	Dignity	Unethical
Gay Marriage	Love	One man and one woman
	Right to marry	Moral and religious
Government:	Government	Washington
Gun:	Gun control	Gun safety
Health:	Healthcare choice	Right to choose
Immigration:	Undocumented	Illegal alien
	Innocent	Outcast
	Unarmed	Obscure
Jobs:	Business	Workers
	Mortgages	Families
	Products	Votes
Lawyer:	Trial lawyer	Personal injury lawyer
Private Property:	Right to use Eminent	Domain Respect ownership
Religion	God and country	Separation of church and state
Tax:	Estate tax	Death tax
	Tax reform	Tax simplification
	Higher tax	Lower tax
Trade:	Foreign Trade	International Trade

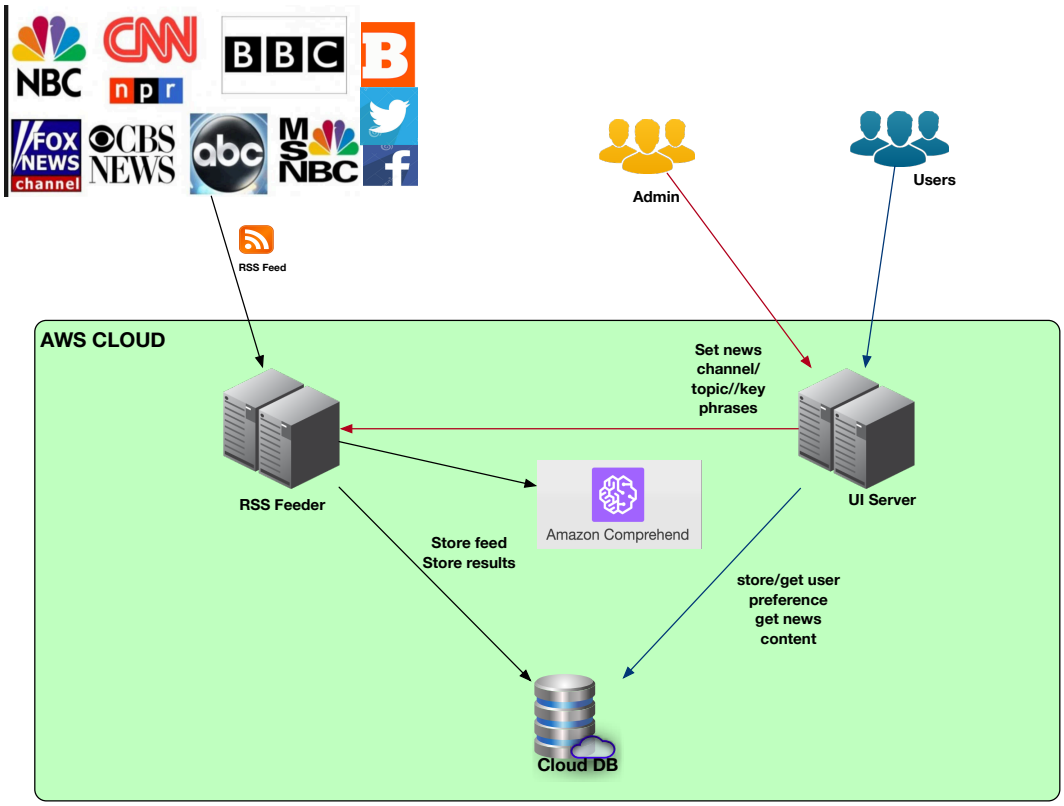
e. Amazon comprehend:
Amazon Comprehend is a natural language processing (NLP) service that uses machine learning to find insights and relationships in text. The service identifies the language of the text; extracts key phrases, places, people, brands, or events; understands how positive or negative the text is; analyzes text using tokenization and parts of speech; and automatically organizes a collection of text files by topic. Using these APIs once can analyze text and apply the results in a wide range of applications including voice of customer analysis, intelligent document search, and content personalization for web applications.

f. UI Server:
This is the presentation tier that serves the user interface to the clients. There are multiple actors who will be interacting with the system. The UI interacts with the RSS feeder to store data. It also interacts with the cloud DB, as a read only mode, to get the user necessary content relevant to his choice.

g. Actors:
The actors are responsible for interacting with the system. The system also plays as an actor. The actors are:

- Users: They are the end clients who shall be interacting with the system. They will select the type of news, the topic and the news channel they are interested in. once they select their preference, the news content will be delivered to them as per their preference. For example, users can select multiple media sources having topic "midterms election" that has positive sentiment and is biased towards liberal ideas.
- Administrator: The administrator has the ability to define the topic, content, the news channel from where the news would be pulled in; they can filter content and can also restrict/unrestricted topics in the interest. They can define the keywords/phrases and define buckets for those. They would add weight to the news media and can define number of buckets where they fall into; they can also mark news item as false/fake based on their individual analysis.
- System as Actor: The system, based on the configuration, pulls the RSS feed from the online news media. They can also pull certain topic news, certain author articles in summary form or in detailed form.

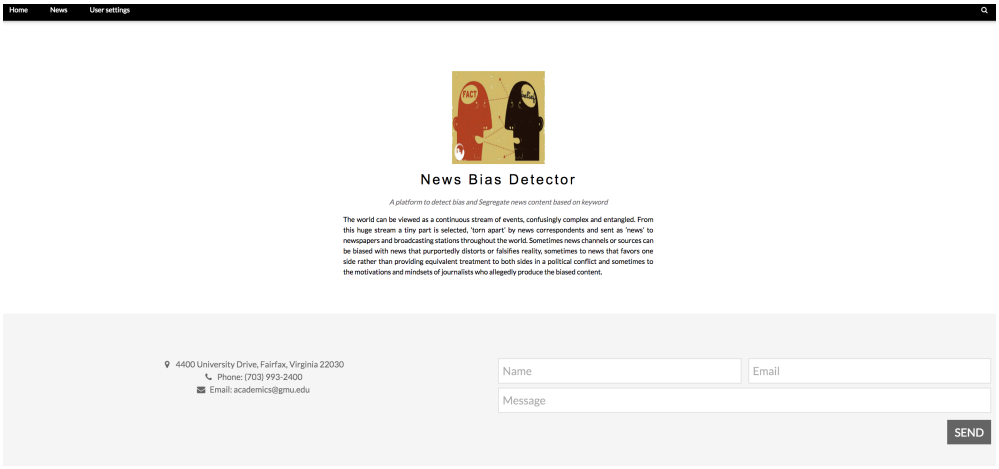
h. System diagram:



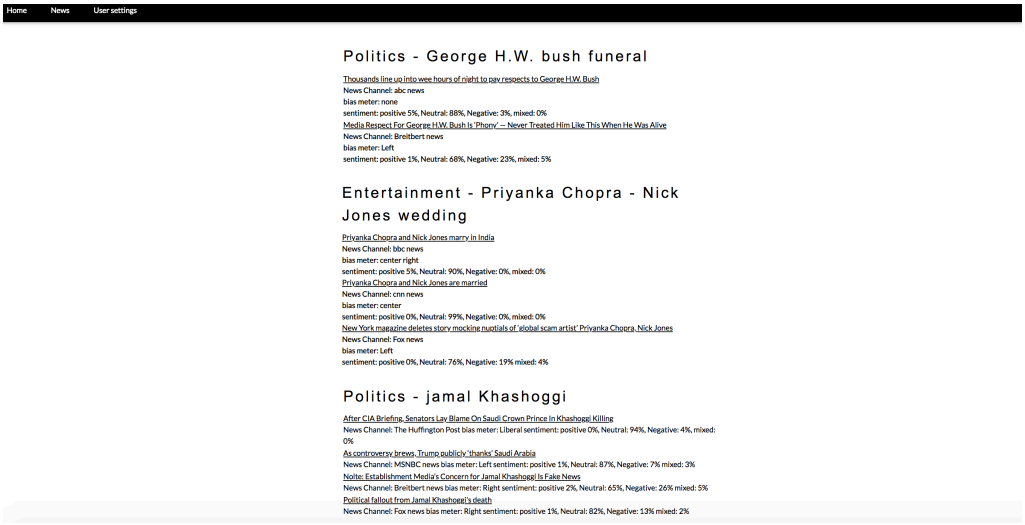
screenshots:

the project will have two types of users –

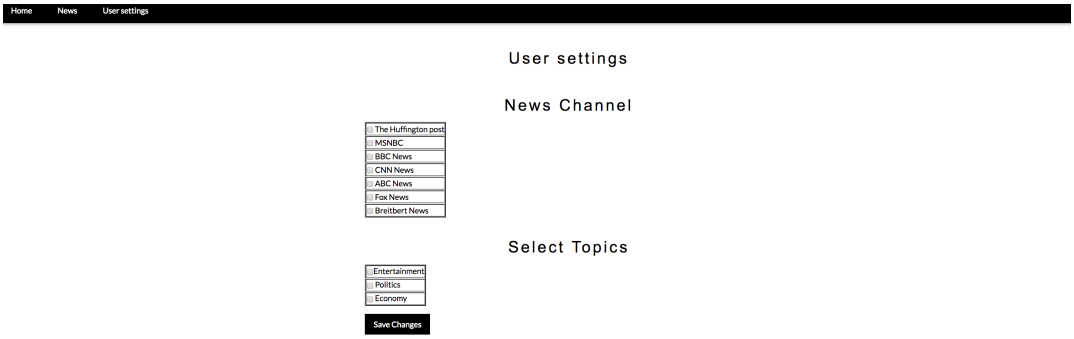
1. Customer who is availing the news service.
    a. Home screen



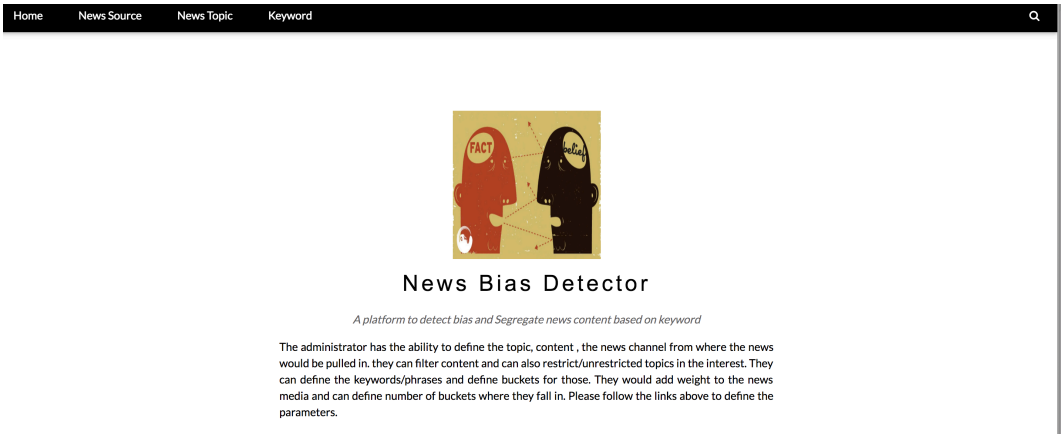
    b. News screen



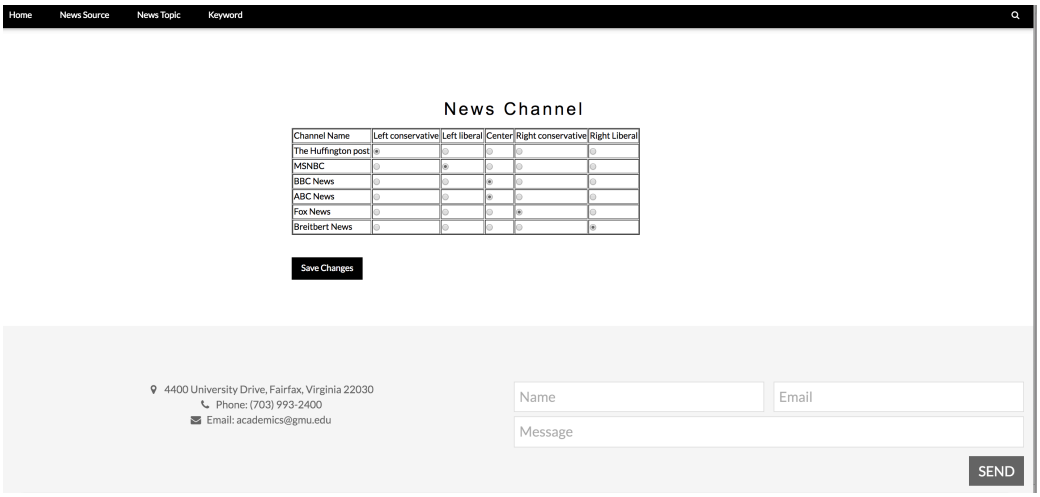
c. User setting screen



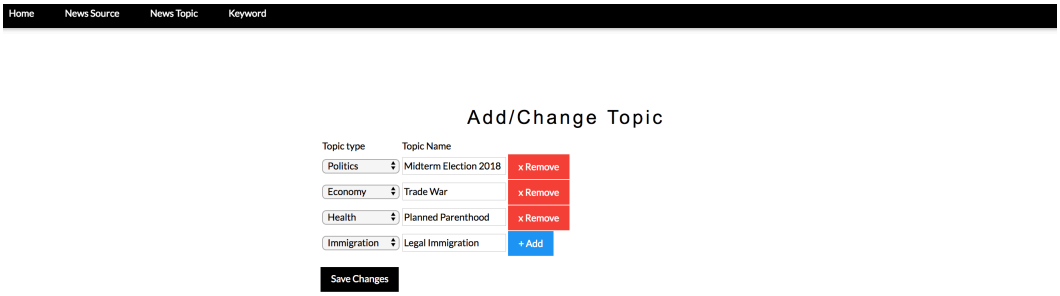
2. Admin who configure the system and the news sources and the topics.
   a. Home screen



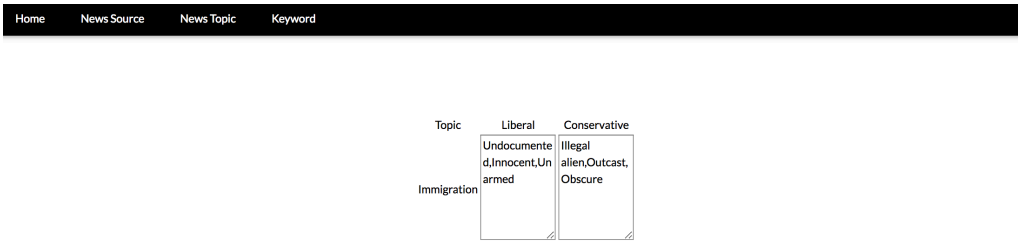
   b. News source configuration screen



c. News topic screen



d. Keyword screen



V. CONCLUSION AND FUTURE WORK

In this paper we have produced a model that will help detecting the bias. The concept is poewerful in terms of present divided world and will help eliminate news which tends to spread negative sentiment. This also can be the founding stone for detecting fake news because fake news essentially starts with sentiment opposite to popular and established news media and/or social media channels. We have produced UI interfaces, which is used to configure the parameters based on learning. The data , in this everchanging world, will change often and will mature by course of time. Our small attempt to capture data pertaining to present scenario can be extended further to the following –

- Expand key word cloud – each keyword can be significant and can determine bias/sentiment. It is very important picking up the right keyword. The algorithm has the opportunity to improve for picking up the right keywork corresponding to a topic.
- Add more sources – the paper limits the sources to news channel but this can be expanded to other social channels like Facebook and twitter. This will expand the sentiment universe and will help ore detecting biases across general public and news channels and help detect fake news.
- Configurability and Implementation – The project is theoretical at this point time but has the potential to route in implementation phase. There could be more functional level and system level configuration that can be added for better performance.
- Publish application – The application can be published externally and be consumable by different sects of people who are interested to read specific type of news.

VI. ACKNOWLEDGEMENT

The team would like to thank Dr. Larry Kerschberg for his guidance, support and valuable suggestion for this project.

VII. REFERENCE LIST

[1] Journals, "Media Bias", Web: https://journals.sagepub.com/doi/pdf/10.1177/002234337100800303, 13th November 2018
[2] Theodysseyonline, "Media Bias", Web: https://www.theodysseyonline.com/why-it-is-important-to-be-aware-of-media-biases, 14th November 2018
[3] Communication cache, "Media Bias", Web: http://www.communicationcache.com/uploads/1/0/8/8/10887248/framing_bias-_media_in_the_distribution_of_power.pdf, 14th November 2018
[4] Wikipedia, "Media Bias", Web: https://en.wikipedia.org/wiki/Media_bias, 15th November 2018
[5] ISTE, "Related works", Web: https://www.iste.org/explore/articleDetail?articleid=916, 20th November 2018
[6] Media Bias Chart: Version 4.0 - ad fontes media Web: https://www.adfontesmedia.com/
[7] Where do news sources fall on the political bias spectrum? - "Fake News," Lies and Propaganda: How to Sort Fact from Fiction - Research Guides at University of Michigan Library Web: https://guides.lib.umich.edu/c.php?g=637508&p=4462444
[8] How to Identify Liberal Media Bias -- Media Research Center -- By: Brent Baker with Forward By: L. Brent Bozell III Web: http://archive.mrc.org/books/identifybias.asp

[9] Journalism, fake news and disinformation: a model course for journalism educators and trainers worldwide; 2018 Web: http://unesdoc.unesco.org/images/0026/002642/264240e.pdf
[10] Fake News Detection with Deep Diffusive Network Model Web: https://arxiv.org/pdf/1805.08751.pdf
[11] Detecting Fake News using Sentiment Analysis Web: https://devpost.com/software/detecting-fake-news-using-sentiment-analysis
[12] Fighting **Fake News** Content with AI, Sentiment Analysis Tech and Machine Learning Web: https://martechseries.com/mts-insights/staff-writers/fighting-fake-news-content-with-ai-sentiment-analysis-tech-and-machine-learning/