# Differences in Health News from Reliable and Unreliable Media Sameer Dhoju<sup>1</sup>, Md Main Uddin Rony<sup>1</sup>, Muhammad Ashad Kabir<sup>2</sup>, Naeemul Hassan<sup>1</sup>

<sup>1</sup>The University of Mississippi, <sup>2</sup>Charles Sturt University

## Motivation

Doctor Blows Whistle on Flu Shot: 'It's Designed to Spread Cancer' April 26, 2018 by Edward Morga



tatement and reveal that flu vaccines have been laced with "cano causing ingredients



- More than 50% of the top-20 Facebook stories containing "cancer" in headline were False.
- Fake news about vaccine caused measles outbreak in Europe.

## **Deluge of misleading health news over social media**

- Continuously produced and propagated by unreliable outlets.
- Reach to a broader audiences through Social networking sites.

## Health misinformation can be critical

- Erroneous health news can cause hazardous health condition.
- Spoil the credibility of the health-care providers & medicines.

## Health misinformation is a relatively unexplored area

- Lack of reliable entities to debunk health misinformation.
- Very few computational approaches with limited success.



- Develop a health-oriented news corpus.
- Analyze the corpus to identify discriminating features.
- Build a classifier to distinguish unreliable media sourced health news from reliable news.





- Media outlets use longer headline to get more attention <sup>3</sup>.
- Unreliable outlets (**40.03%**) practice more clickbait than
- Unreliable outlets use demonstrative adjective and numbers significantly more than the reliable outlets.

Numbers -

**Demonstrative Adjectives** -



References

**Distribution of clickbait patterns** 

# **Topical Analysis**

Used Google Cloud NLP & Latent Dirichlet Allocation (k=3). Reliable & unreliable outlets cover different health topics.

new cancer

day Skin like infection ⊐ @ Virus ↓

For Reliable News Articles

![](_page_0_Picture_45.jpeg)

For Unreliable News Articles

## **Topic Modeling (RT denotes Reliable and UT denotes Unreliable Topic)**

# Semantic Analysis

Use of quotations & links indicates credibility of an article <sup>1,2</sup>. Reliable Sources 0.175 Unreliable Sources 0.08 0.150 0.125 0.06 0.100 ŏ <sub>0.075</sub> ٥.04 🗅 0.050 0.02 0.025 0.000

Average number of Direct Quotations in a Source

**Distribution of avg. number of quotation/link per article** 

![](_page_0_Figure_55.jpeg)

3. Chris Breaux. (accessed September 28, 2018). "You'll Never Guess How Chartbeat's Data Scientists Came Up With the Single Greatest Headline". https://tinyurl.com/nleq7ph

# Charles Sturt University

Labels	Precision	Recall	<b>F-1</b>
Unreliable	0.94	0.92	0.93
Reliable	0.96	0.97	0.97
Macro-Avg	0.95	0.95	0.95
Unreliable	0.76	0.47	0.58
Reliable	0.78	0.93	0.85
Macro-Avg	0.77	0.70	0.72
Unreliable	0.95	0.93	0.94
Reliable	0.97	0.97	0.97
Macro-Avg	0.96	0.95	0.96

**Classification Report for different combination of features** 

1. De Maeyer, J. (2012). The journalistic hyperlink: Prescriptive discourses about linking in online news. Journalism *Practice*, *6*(5-6), 692-701.

2. Sundar, S. S. (1998). Effect of source attribution on perception of online news stories. Journalism & Mass Communication *Quarterly*, *75*(1), 55-68.