

## Uncovering the multifaceted nature of fake news: Latent dirichlet allocation topic modelling approach

Abishethvarman V.<sup>1\*</sup>, Banujan K.<sup>1</sup>, Nirubikaa R.<sup>2</sup>, Kumara B.T.G.S.<sup>1</sup>

<sup>1</sup>*Department of Computing and Information Systems, Sabaragamuwa University of Sri Lanka, Belihuloya, Sri Lanka*

<sup>2</sup>*Department of Software Engineering, Sabaragamuwa University of Sri Lanka, Belihuloya, Sri Lanka*

Disseminating fake news is a critical challenge in the digital information world. Misleading or fabricated information, masquerading as reliable news, has impacted public opinion and information integrity. Fake news is subjective to its context. This research aims to uncover diverse thematic structures within spanning topics of fake news through Latent Dirichlet Allocation (LDA) and visualizers. The well-established fake news labeled dataset was taken and then we followed the preprocessing techniques such as text normalization, punctuation, stop words, and custom word removal. We extracted features through CountVectorizer and moved for the topic modeling. The LDA model was employed to find ten topics. As a result, topics encompass keywords which are diverse themes, from everyday discourse to electoral dynamics, governance, and global affairs. Some topics underscore the role of news outlets in propagating misinformation, emphasizing the need for source scrutiny. Continued iterations made this more accurate in approach. Additionally, this research delves into political scenarios, gender issues, investigations, law enforcement, and political campaigns, revealing the multifaceted nature of fake news challenges as topics analyzed through the LDA visualizer. Visualizer helps to differentiate the nuanced dimensions of fake news between each topic. Each topic was analyzed through its semantic value in the fake news. The results emphasize the commonly used words in the fake news corpus. That pinpointed the triggering words in fake news, distinguishing them from reliable news through topic modeling and analysis. This methodology can be extended to user profiling based on the extracted keywords, into characteristics of individuals interacting with online content.

**Key words:** Fake News, Latent Dirichlet Allocation (LDA), Topic Modelling, LDAvisualizer

\*Corresponding author: abishethvarman@gmail.com