Sinhala Language Fake News Detection in Social Media Using Autoencoder-Based Method

Rahul Adihetti[1*], Shantha Jayalal[2]

[1]*Department of Industrial Management, University of Kelaniya, Sri Lanka,*
rahulkeshara@gmail.com

[2]*Department of Industrial Management, University of Kelaniya, Sri Lanka, shantha@kln.ac.lk*

The spread of fake news in the social media has grown significantly over the past few years. According to the New York Times, fake news is defined as "made-up articles meant to deceive." Additionally, the way they are released is almost identical to that of conventional news organizations. The issue is that a significant number of news outlets outside the major and reliable ones are disseminating unreliable information. This problem is exacerbated by the ease with which anything can be published from anywhere on well-known social networking and social media platforms. People can use this to their advantage by disseminating any type of message on various social networking sites to accomplish their objectives. In the Sri Lankan context, content posted in Sinhala greatly impacts fake news in Sri Lanka. Because utilizing the Sinhala language to describe emotions and feelings makes it easier to connect with Sinhala-speaking people than using content that has been published in other languages, like English. The use of Sinhala on social media has grown over the past few years. Additionally, as the use of the Sinhala language expanded, so did the number of occurrences of fake news. Based on the literature, approaches to identifying fake news depend on the features of the news content. Therefore, this research proposed an autoencoder- based method for Sinhala fake news detection, which is an unsupervised method. The method uses Text, User, Propagation, and Image features from the news content. And also, this research found the best feature combination to detect Sinhala language fake news content, which is a combination of Text, User, and Image features. The method gained an accuracy of 98% and 88% in Precision, Recall, and F1 Score by outperforming other existing anomaly detection methods. The main stakeholder of this study was fact-checking organizations in Sri Lanka.

Keywords: *Sinhala language, fake news detection, social media, unsupervised learning, neural networks*