"The Truth Will Come Out": Defending Against the Viral Spread of Misinformation by Characterizing User Responses to Fake News

Jethwani-Keyser, Taj (School: The Bronx High School of Science)

This study applies natural language processing to fake news comment sections as a novel framework for efficient, real-time detection. Online social networks are essential tools to communicate and spread information, making it crucial to identify and eliminate fake news posts on these platforms before they achieve a viral spread. However, most automated detection systems examine solely the text content of potential fake news and ignore additional context. This approach is highly susceptible to false negatives since classifiers will return the same result on the same batch of text, endangering governments, public health, and societal trust. In this study, fact-checking efforts by concerned citizens who comment on misinformation posts are explored as a potential method to address this weakness. A text-based deep learning model (F1-score = 0.704) is developed to classify user responses to online content. Then, in a big-data analysis of responses to 400,118 Tweets related to the COVID-19 vaccine, it is observed that users fact-check false content nearly three times as frequently (%-difference = 69.8%) compared with reliable news. Results suggest a positive relationship between the number of counter-misinformation replies and the probability that content contains falsehoods. An adaptive fake news algorithm is proposed that analyzes replies to online posts and improves in precision as they receive attention. This research contributes to a growing field dedicated to examining the role of the crowd in fighting fake news, and opens avenues to further leverage this resource in the future.