

Review of: "Methods of Identifying Fake News in Social Networks"

Bhuvanesh Singh¹

¹ Independent researcher

Potential competing interests: No potential competing interests to declare.

The authors have experimented to classify the text as fake or not using a calculated weightage method with word tokenization. There are multiple mismatches in the information provided about the dataset and the method. Authors should update this information for better clarity of the paper.

1. Labeling: The authors say they collected 1400 texts from various platforms. But how did they label them? Platforms like Twitter and Instagram do not have labels as fake or real. What strategy was taken to label the dataset as fake/real? What was the review process for labeling?

2. There are a couple of mismatches in information. Authors are requested to update/clarify.

- In the abstract, authors say "*News collected during August to November 2024*", but in the dataset section, they say "*September 2024 to October 2024 were analysed*"
- Authors say "*One thousand four hundred news stories*", but in figure 1, the total is 1389 only.
- Authors say "*We use the training set to train the SVM model*", but no information is provided about the SVM model anywhere in the code and content of the paper. Please provide information on how and at which stage the SVM model was used. Please share the code as well.

3. Please clarify the "news on sites"? What are they? Please mention.

4. Please clarify "*70-80% for training and 20-30% for testing*", what percentages were selected?

5. Please share the results of the training and test sets separately.

6. Please share more information on other metrics like Precision, Recall, F1-Score.

7. Please explain the reasoning behind "*words (50%), language (25%), and language (25%)*", from where is this percentage threshold coming?

