Explainable Fake-News Detection and Personalized Credible Recommendation via GraphML

Mahnoor Sheikh

1. Problem Description

The proliferation of fake news on critical topics remains a persistent problem on the web. False information is often used as a means to spread hysteria and panic among gullible readers who seldom fact-check. Manual verification itself is a cumbersome task requiring extensive research and critical evaluation, making it time-consuming and impractical for everyday readers.

To address this issue, I will develop a system that allows users to input the URL of a news article and receive a credibility assessment. The algorithm will (i) classify the article as 'real' or 'fake', (ii) provide an explanation for the decision, and (iii) recommend three relevant and safe alternatives; selected based on source reliability, content similarity, and user network popularity. This approach will make fact-checking time-efficient, transparent, and accessible.

2. Preliminary Plan

I plan to use the FakeNewsNet dataset containing articles from PolitiFact and GossipCop. PolitiFact strives to publish accurate political news, whereas GossipCop focuses on fact-checked celebrity reporting. The dataset contains the following fields:

```
"id": Unique identifier for each news
"url": Url of the article from web that
    published that news
"title": Title of the news article
"tweet_ids": Tweet ids (list separated by
    tab) of tweets sharing the news
```

The baseline model will employ TF-IDF scores with logistic regression on text data for binary classification. To scale this up, XGBoost will be trained on article embeddings or TF-IDF scores. Additionally, LightGCN will be trained on user-article interactions for graph-based recommendations.

The novel approach will experiment with DistilBERT for the binary classification task and concatenate BERT embeddings with LightGCN or GraphSAGE embeddings to create a hybrid fake news detection and article recommendation system.

For explainability, SHAP will be applied to text models, and GNNExplainer to graph-based recommendations.

3. Reference Papers

Golovin et al. (2025) used knowledge graphs with graph attention networks to improve social network recommendations for fake news detection. Malik et al. (2024) built ensemble GNNs by combining text and user engagement, showing stronger performance than single models. Shu et al. (2019) proposed dEFEND, an explainable model using attention over text and user interactions. These works align with my plan to combine BERT transformer embeddings with GNN-based methods and add explainability.

References

Golovin, A., Zhukova, N., Delhibabu, R., and Subbotin, A. Improving recommender systems for fake news detection in social networks with knowledge graphs and graph attention networks. *Mathematics*, 13(6):1011, 2025. ISSN 2227-7390. doi: 10.3390/math13061011. URL https://www.mdpi.com/2227-7390/13/6/1011.

Malik, A., Behera, D. K., Hota, J., and Swain, A. R. Ensemble graph neural networks for fake news detection using user engagement and text features. *Results in Engineering*, 24:103081, 2024. ISSN 2590-1230. doi: 10.1016/j.rineng.2024.103081. URL https://www.sciencedirect.com/science/article/pii/S2590123024013367.

Shu, K., Cui, L., Wang, S., Lee, D., and Liu, H. dEFEND: Explainable fake news detection. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '19, pp. 395–405, Anchorage, AK, USA, 2019. Association for Computing Machinery. doi: 10.1145/3292500.3330935. URL https://doi.org/10.1145/3292500.3330935.