

Assignment 2

Course Title: Research Methodology in Computer Science

Course Code: AI800

Weighting: 30%

Lecturer: Dr. Mohammed Anbar

Hand-out Date: 28 February 2026

Hand-in Date: 2 April 2026

Submission Method: Online Submission via E-Learning

Research Title

Improving Fake News Detection by Combining Machine Learning and Social Network Features

1. Problem Statement

Over the past decade, the rise of social media platforms, including Twitter, Facebook, and TikTok, has fundamentally transformed the way people consume and share information. News and updates now reach millions of users within minutes, while this connectivity facilitates rapid communication. However, this same speed and reach have also created the perfect environment for the spread of fake news, false or misleading information deliberately crafted to manipulate public opinion or attract attention. When such misinformation circulates widely, it can cause substantial harm, damaging reputations, influencing electoral outcomes, and inciting public panic during critical events, such as health crises. As society becomes increasingly dependent on digital platforms for news, addressing the challenges posed by fake news has become an urgent priority.

Numerous researchers have explored ways to detect false information early. Machine learning (ML) and deep learning are the two main approaches that have shown remarkable progress in automatically identifying misleading content. For instance, Ahmed et al. (2017) show that algorithms like Support Vector Machines (SVM) and Naive Bayes can classify news articles based on textual features such as wording, frequency of terms, and writing style. Similarly, Kaliyar et al. (2021) developed the *FNDNet* model using deep convolutional neural networks that achieved high accuracy on benchmark datasets. These studies suggest that computers can be trained to differentiate real news from fake by learning from patterns within language itself.

However, one main limitation of these content-based approaches is that they rely only on the text itself. They often overlook the broader social context in which fake news circulates. Research by Shu et al. (2017) and Monti et al. (2019) shows that the way people share, repost, comment, or interact with news can reveal a lot about its credibility. When a story spreads unusually fast or mainly within certain online communities, it often shows traits typical of misinformation. For that reason, combining content analysis with the study of social network behavior could result in stronger and more dependable detection models.

Another challenge lies in the datasets used to train and test fake news models. For example, Wang (2017) developed the *LIAR* dataset, which quickly became a common benchmark for testing how well algorithms perform. However, this dataset mainly contains short political statements, so models trained on it don't always perform well in other areas such as health, business, or environmental news. In addition, some of the most useful social media data, like user interactions and sharing behavior, are often off limits to researchers because of privacy restrictions. As a result, it becomes difficult to design real world systems that reflect both the textual and contextual aspects of fake news.

Given these limitations, a clear research gap still exists. Most current models either depend mostly on text features or rely on complex data that's often difficult to access. What hasn't been explored enough is how to bring together linguistic features, the actual content of the news, and contextual features such as social interactions and sharing patterns into one hybrid model that's both practical and adaptable across different domains. Tackling this issue is important for developing more robust and fair detection systems that can combat misinformation on a wider scale.

This study will focus on developing a hybrid fake news detection framework that uses both text information and social context features. By bringing these two perspectives together, the proposed method aims to improve classification accuracy, enhance its generalization across different news domains, and capture the constantly changing nature of information flow online. In the long run, this research seeks to contribute to efforts against misinformation and help social platforms strengthen public trust and maintain their integrity in digital spaces.

2. Research Objectives

Main Objective:

To develop an improved hybrid fake news detection model that integrates machine learning text analysis with social context information to increase accuracy and adaptability.

Specific Objectives

- 1. To review and synthesize previous studies on content-based, context-based, and hybrid approaches to fake news detection.**

This objective builds on the work of Shu et al. (2017) and Zhou and Zafarani (2020), who offer overviews of fake news detection methods from both content and social perspectives. By critically examining these and other studies, this research will identify how the field has evolved from traditional text-focused machine learning techniques using handcrafted features, to more integrated frameworks that include user behavior and network patterns. Reviewing this literature will help pinpoint methodological strengths, persistent challenges, and the theoretical gaps justifying a new hybrid approach.

- 2. To analyze and compare existing benchmark datasets used for training and testing fake news detection models.**

Building on Wang's (2017) LIAR dataset and other commonly used corpora, this objective seeks to assess their adequacy and representativeness across different subject areas. It will involve examining dataset composition, labeling methods, and domain coverage to determine which linguistic, semantic, and social features are

most useful for distinguishing truthful from deceptive news. The analysis will also highlight the limitations of existing benchmark data, such as restricted topic diversity or the scarcity of labeled social context information, as noted in Monti et al. (2019).

3. **To design a hybrid machine learning model that combines linguistic and social context features for comprehensive fake news detection.**

Drawing from evidence in Ahmed et al. (2017) and Kaliyar et al. (2021), this objective will guide the design of a hybrid framework that merges traditional text analysis (n-gram features, sentiment indicators, and writing style) with social network characteristics (user interactions, propagation patterns, and engagement metrics). The goal is to leverage the high classification accuracy of deep learning models such as FNDNet while enriching them with context-level signals inspired by graph-based learning approaches proposed by Monti et al. (2019).

4. **To implement and experimentally evaluate the proposed hybrid model using suitable datasets.**

The implementation phase will test the model on benchmark datasets like LIAR and other open-source data that capture both textual and contextual aspects of social media activity. Where possible, public posts or anonymized datasets will be used to ensure ethical compliance. The evaluation will focus not only on accuracy metrics but also on computational efficiency and the model's adaptability across different domains, as earlier studies emphasize trade-offs between performance and scalability.

5. **To compare the proposed hybrid model with conventional and deep learning approaches to assess improvement and generalization**

Following insights from earlier comparative research—such as Ahmed et al. (2017), who contrasted Naïve Bayes and SVM models, and Kaliyar et al. (2021), who evaluated CNN architectures—this objective involves benchmarking the hybrid model against individual models. The results will be measured using accuracy, precision, recall, and F1-score. This comparison will determine whether the integration of social context data produces statistically significant gains over purely text-based or graph-based approaches, and whether the model can generalize effectively beyond the datasets it was trained on.

6. **To discuss future directions and practical implications for deploying hybrid fake news detection systems**

Based on findings from studies like Zhou and Zafarani (2020), which outline research opportunities and challenges, this final objective will explore practical aspects of

applying the proposed model within actual social media environments. Topics such as data privacy, user trust, interpretability, and ethical considerations in algorithmic moderation will be addressed to ensure the proposed solution is both effective and socially responsible.

To build real trust in automated detection systems, it's important to handle user data responsibly and reduce any bias that might influence the model's decisions.

3. Expected Contribution

This research is expected to make meaningful contributions to combating misinformation in online spaces. By integrating lessons from previous studies such as Monti et al. (2019) and Kaliyar et al. (2021), it aims to create a practical hybrid system that merges text analysis with social behavior modeling. The results could aid fact-checkers and media analysts in identifying false information before it spreads widely. With academic knowledge, the project emphasizes social responsibility, addressing issues of user privacy, data fairness, and algorithmic transparency to ensure that fake news detection tools enhance public trust instead of undermining it. The proposed framework may also inspire similar solutions in other domains like health communication, education, and digital safety.

4. References

- Ahmed, H., Traore, I., & Saad, S. (2017). *Detection of Online Fake News Using N-gram Analysis and Machine Learning Techniques*.
- Kaliyar, R., Goswami, A., & Narang, P. (2021). *FNDNet: A Deep Convolutional Neural Network for Fake News Detection*.
- Monti, F., Frasca, F., Eynard, D., Mannion, D., & Bronstein, M. (2019). *Fake News Detection on Social Media Using Geometric Deep Learning*.
- Shu, K., Sliva, A., Wang, S., Tang, J., & Liu, H. (2017). *Fake News Detection on Social Media: A Data Mining Perspective*.
- Wang, W. Y. (2017). *'Liar, Liar Pants on Fire': A New Benchmark Dataset for Fake News Detection*.
- Zhou, X., & Zafarani, R. (2020). *A Survey of Fake News: Fundamental Theories, Detection Methods, and Opportunities*.