

# **Fake news detection in online social media**

## **Problem Statement**

Social media for news consumption is a double-edged sword. On the one hand, its low cost, easy access, and rapid dissemination of information lead people to seek out and consume news from social media. On the other hand, it enables the wide spread of “fake news”, i.e., low quality news with intentionally false information. The extensive spread of fake news has the potential for extremely negative impacts on individuals and society. Therefore, fake news detection on social media has recently become an emerging research that is attracting tremendous attention. Fake news detection on social media presents unique characteristics and challenges that make existing detection algorithms from traditional news media ineffective or not applicable. First, fake news is intentionally written to mislead readers to believe false information, which makes it difficult and nontrivial to detect based on news content; therefore, we need to include auxiliary information, such as user social engagements on social media, to help make a determination. Second, exploiting this auxiliary information is challenging in and of itself as users’ social engagements with fake news produce data that is big, incomplete, unstructured, and noisy.

## **Background**

Detecting fake news on social media poses several new and challenging research problems. Though fake news itself is not a new problem—nations or groups have been using the news media to execute propaganda or influence operations for centuries—the rise of web-generated news on social media makes fake news a more powerful force that challenges traditional journalistic norms. There are several characteristics of this problem that make it uniquely challenging for automated detection. First, fake news is intentionally written to mislead readers, which makes it nontrivial to detect simply based on news content. The content of fake news is rather diverse in terms of topics, styles and media platforms, and fake news attempts to distort truth with diverse linguistic styles while simultaneously mocking true news. For example, fake news may cite true evidence within the in-correct context to support a non-factual claim. Thus, existing hand-crafted and data-specific textual features are generally not sufficient for fake news detection. Other auxiliary information must also be applied to improve detection, such as knowledge base and user social engagements. Second, exploiting this auxiliary information actually leads to another critical challenge: the quality of the data itself. Fake news is usually related to newly emerging, time-critical events, which may not have been properly verified by existing knowledge bases due to the lack of corroborating evidence or claims. In addition, users’ social engagements with fake news produce data that is big, incomplete, unstructured, and noisy. Effective methods to differentiate credible users, extract useful post features and exploit network interactions are an open area of research and need further investigations.

## **Methodology**

### **Step 1: Feature Extraction**

News content features describe the meta information related to a piece of news. A list of representative news content attributes are listed below:

- Source: Author or publisher of the news article
- Headline: Short title text that aims to catch the attention of readers and describes the main topic of the article
- Body Text: Main text that elaborates the details of the news story; there is usually a major claim that is specifically highlighted and that shapes the angle of the publisher
- Image/Video: Part of the body content of a news article that provides visual cues to frame the story.

Based on these raw content attributes, different kinds of feature representations can be built to extract discriminative characteristics of fake news. Typically, the news content we are looking at will mostly be linguistic-based and visual-based.

## **Step2: Model Construction**

Since fake news attempts to spread false claims in news content, the most straightforward means of detecting it is to check the truthfulness of major claims in a news article to decide the news veracity. Knowledge-based approaches aim to use external sources to fact-check proposed claims in news content. The goal of fact-checking is to assign a truth value to a claim in a particular context. Fact-checking has attracted increasing attention, and many efforts have been made to develop a feasible automated fact-checking system. Existing fact-checking approaches can be categorized as expert-oriented, crowdsourcing-oriented, and computational-oriented.

## **Experimental Design**

**Datasets:** Online news can be collected from different sources, such as news agency homepages, search engines, and social media websites. However, manually determining the veracity of news is a challenging task, usually requiring annotators with domain expertise who performs careful analysis of claims and additional evidence, context, and reports from authoritative sources. Generally, news data with annotations can be gathered in the following ways: Expert journalists, Fact-checking websites, Industry detectors, and Crowd-sourced workers

**Evaluation Metrics:** evaluate the performance of algorithms for fake news detection problem, various evaluation metrics have been used. In this subsection, we review the most widely used metrics for fake news detection. Most existing approaches consider the fake news problem as a classification problem that predicts whether a news article is fake or not:

- True Positive (TP): when predicted fake news pieces are actually annotated as fake news;
- True Negative (TN): when predicted true news pieces are actually annotated as true news;
- False Negative (FN): when predicted true news pieces are actually annotated as fake news;
- False Positive (FP): when predicted fake news pieces are actually annotated as true news.

By formulating this as a classification problem, we can define

following metrics,

1. Precision =  $\frac{TP}{TP + FP}$
2. Recall =  $\frac{TP}{TP + FN}$
3. F1 =  $\frac{2 \cdot \text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}}$
4. Accuracy =  $\frac{TP + TN}{TP + TN + FP + FN}$

These metrics are commonly used in the machine learning community and enable us to evaluate the performance of a classifier from different perspectives. Specifically, accuracy measures the similarity between predicted fake news and real fake news.