

Misinformation in News Detection Using NLP

COE 379L Project Proposal

Desheng Liu
Vrinda Pandey

Proposal

In today's world of digital information and the increasing use of generative AI, the spread of misinformation—especially in the political domain—has become a serious societal challenge. False or misleading news stories can influence public opinion, destroy trust, disrupt democratic processes, or lead to defamation. Social media platforms and online news companies that have millions of followings and user base, which make it easy for such content to go viral, often without fact-checking or source verification.

This project aims to address the problem of political misinformation by building a misinformation news detection system using natural language processing (NLP). The goal is to train a model that can classify news articles as either real or false (*"fake" as referenced by the dataset*), based on their text content and genre of news. By analyzing patterns in language, syntax, and word usage, the model will learn to identify characteristics that often distinguish false news.

Datasets

I plan to use a Kaggle dataset, specifically the *"misinformation-fake-news-text-dataset"*, which linked here:

<https://www.kaggle.com/datasets/stevenpeutz/misinformation-fake-news-text-dataset-79k>

This dataset is arguably reputable, as there is a 10.00 usability rating on Kaggle. There are roughly 40,000 data points for each class, which is arguably enough to produce credible and successful results.

High-Level Methods. Techniques, Technologies

This project will explore multiple natural language processing techniques to determine the most effective approach for detecting fake news. I plan to experiment with both classical machine learning models and transformer-based models, comparing their performance to select the best-performing approach. For classical models, the text will be preprocessed using standard techniques that we learn in class, along with additional new preprocessing techniques I will research on. These include lowercasing, punctuation removal, and stopword removal. The cleaned text will be converted into numerical features using TF-IDF (learned in an additional data science class) vectorization and used to train models such as logistic regression, naive Bayes, and random forest.

I also plan to fine-tune a pre-trained transformer model (looking into specifically DistilBERT from Hugging Face). Furthermore, I will evaluate the model on both before and after fine-tuning to see the results.

I will look at results by the metrics we learned in class as usual – accuracy, precision, recall, and F1-score. Furthermore, I will also use the confusion matrix and ROC curves to visually analyze classification performance. Ultimately, these metrics will be the factors in choosing the best model for the misinformation

news detection system.

Project Deliverables

The primary deliverables for this project will include all components necessary to demonstrate model development, evaluation, and reproducibility. These will be:

- The dataset used for training and testing
- A complete preprocessing and modeling pipeline in code
- Trained models: classical and transformer-based
- Evaluation results (accuracy, precision, recall, F1-score, confusion matrix, ROC curve)
- A written final report summarizing the following:
 - Intro / Project Statement
 - Data Sources / Technologies Used
 - Methods employed
 - Results / References
- A video presentation (maximum of 10 minutes) explaining the following:
 - The written report
 - Primary aspects of the work
- The GitHub repository (with link) containing all project files, proposal, and final report