

PROJECT PROPOSAL - CAPSTONE 3

Problem Statement Worksheet (Hypothesis Formation)

Develop an automated NLP-based classification system that identifies whether a news article is fake or real within seconds, reducing manual review time from multiple days to near real-time, while achieving an F1-score of at least 85% on a labeled news dataset.



1 Context

The rapid spread of online news has increased the circulation of misinformation, making it difficult for readers and organizations to verify content quickly. Traditional manual fact-checking processes are time-consuming and allow fake news to spread before validation. This project uses Natural Language Processing (NLP) to automatically classify news articles as fake or real, enabling faster verification, improving trust in credible sources, and reducing the impact of false information.

2 Criteria for success

Readers and organizations must be able to trust credible sources and limit the impact of false information.

3 Scope of solution space

Automated NLP-based classification system with at least 85% F1 score has to be developed for fake news detection.

4 Constraints within solution space

- Limited by data quality and labeling
- Domain- and language-specific performance
- Difficulty capturing context and sarcasm
- Trade-off between accuracy and interpretability
- Computational limits for real-time predictions

5 Stakeholders to provide key insight

1. **Readers / General Public:** End-users of news
2. **News Organizations / Journalists:** Content creators & fact-checkers
3. **Fact-Checking Agencies:** Verification & accuracy
4. **Social Media Platforms:** Content distribution
5. **Data Scientists / NLP Engineers:** Model development
6. **Policy Makers / Regulators:** Compliance & trust
7. **Tech/IT Teams:** Deployment & monitoring

6 Key data sources

ISOT Fake News Dataset from Kaggle;

https://www.kaggle.com/datasets/emineyettm/fake-news-detection-datasets?utm_source=chatgpt.com

Deliverables

A GitHub repo containing the work completed for each step of the project, including:

- A slide deck
- A project report