Eklaro: A System that determine the validity of articles

A Capstone Project

Submitted to the Faculty of

National University - College of Computing and Information Technologies

In Partial Fulfillment of the Requirements for the Degree

Bachelor of Science in Information Technology with

Specialization in Mobile and Web Application

BY:

Asada, Timothy John

Hernandez, Rojen

Tool, Jhiro Ramir

Capstone Adviser

September 2025

CHAPTER 1: INTRODUCTION

1.1 Project Context

In the digital age, the rapid spread of misinformation and disinformation has become a major concern. Social media platforms, blogs, and online news outlets have made it easier than

ever to publish and share articles, but not all content is accurate or trustworthy. This has led to confusion among readers, erosion of public trust, and even real-world consequences in areas such as health, politics, and public safety.

Eklaro is a proposed system designed to assess the validity of articles using natural language processing (NLP), machine learning (ML), and fact-checking databases [1], [2]. The system aims to help users—especially students, researchers, and casual readers—identify whether an article is credible, biased, or misleading. By analyzing linguistic patterns, source reliability, and factual consistency, Eklaro will provide a credibility score and supporting insights to guide informed reading.

This project aligns with the global push for media literacy and responsible information consumption, contributing to Sustainable Development Goal (SDG) 16: Peace, Justice, and Strong Institutions, by promoting access to reliable information.

1.2 Objectives of the Study

General Objective

To design and develop Eklaro, a system that determines the validity of articles using machine learning and natural language processing techniques.

Specific Objectives

1. To collect and preprocess datasets of valid and invalid articles from verified sources and misinformation databases.
2. To develop a classification model that evaluates article credibility based on linguistic features, source metadata, and factual consistency [3].
3. To implement a web-based interface where users can input articles and receive a validity score and explanation.
4. To integrate external fact-checking APIs (e.g., Google Fact Check Tools, ClaimReview) for cross-referencing claims [4].
5. To evaluate the system's accuracy, usability, and performance through testing with real-world articles and user feedback.

1.3 Scope and Delimitation

Scope of the Project

This study focuses on building a software-based system that analyzes the validity of textual articles. It includes data collection, model training, system development, and integration with fact-checking APIs. The system will be accessible via a web interface and will support articles in English during the initial phase.

Delimitation of the Project

The system does not cover multimedia content such as videos or images. It is limited to textual articles and may not detect nuanced satire or sarcasm. The accuracy of the system depends on the quality and diversity of training data and the availability of external fact-

checking sources. Real-time validation of breaking news may be limited due to lag in fact-checking updates.

References

[1] V. Pérez-Rosas, B. Kleinberg, A. Lefevre, and R. Mihalcea, “Automatic detection of fake news,” in Proceedings of the 27th International Conference on Computational Linguistics, 2018, pp. 3391–3401.

[2] C. Castillo, M. Mendoza, and B. Poblete, “Information credibility on Twitter,” in Proceedings of the 20th International Conference on World Wide Web, 2011, pp. 675–684.

[3] A. Pandey and A. K. Jain, “A machine learning approach for detecting fake news using natural language processing,” in 2021 11th International Conference on Cloud Computing, Data Science & Engineering (Confluence), IEEE, 2021, pp. 593–598.

[4] J. Li, Y. Yang, and C. Hauff, “Assessing the credibility of online articles using natural language processing and machine learning,” Journal of Information Science, vol. 51, no. 1, pp. 45–60, 2025.