

Misinformation Detection Pipeline

Jose Fco. Olivert Iserte^{1,*}, Hugo Albert Bonet¹, Kexin Jiang Chen¹, David Borregón Sacristán¹,
Diana Yaser Haj¹, Ivan Arcos Bonet¹

(1) Universitat Politècnica de València,
Camino de Vera, s/n. 46022 Valencia.

Abstract

Misinformation during elections has become a significant concern, as it can influence public opinion and undermine democratic processes [1],[2]. In the context of Nigeria's 2023 Presidential Election, UNICC posed the challenge to build an automatic system capable of effectively detecting false claims, hoaxes, and other forms of misinformation spread on Twitter. To address this challenge, a misinformation detection pipeline was developed, consisting of two main components: a fact-checking algorithm and a machine learning model. State-of-the-art NLP techniques, such as BERT-style transformers, were used to train the model on a dataset of Nigerian political tweets. We have achieved F1-Scores close to 0.9 by combining embeddings and classical classifiers. Incorporating information on toxicities and emotions further improved performance. The fine-tuning technique using a combination of CNN and evidence sentences reached an F1-Score of 0.907, emphasizing the importance of providing supporting information. Additionally, we created a fact-checking database and reusable resources for future projects, showcasing the effectiveness of NLP techniques in detecting misinformation. This endeavour offers additional value by establishing an extensive fact-checking database from scratch and generating reusable Python classes and modules, and project documentation that can facilitate the development of analogous pipelines for future misinformation detection projects. As new elections approach, it would be interesting to monitor the spread of misinformation and evaluate the performance of our system in real-time. This will not only help in mitigating the impact of false claims but also contribute to the ongoing research on misinformation detection during elections.

References

- [1] Vosoughi S, Roy D, Aral S. The spread of true and false news online, *Science*. 2018 Mar 9;359(6380):1146-1151. doi: 10.1126/science.aap9559. PMID: 29590045.
- [2] Nakov et al, Automated Fact-Checking for Assisting Human Fact-Checkers, arXiv:2103.07769v2 [cs.AI] 22 May 2021

* Contact email: jfoliise@etsinf.upv.es