

ABSTRAK

Rendahnya literasi digital masyarakat Indonesia menyebabkan sulitnya membedakan informasi valid dan hoax, khususnya berita tentang politik. Penelitian ini bertujuan membangun deteksi hoax teks berita berbahasa indonesia dengan metodologi Cross Industry Standard Process for Data Mining (CRISP-DM). Dataset diperoleh melalui proses web scraping dari situs resmi Sekretariat Kabinet Republik Indonesia (setkab.go.id). Pengujian dilakukan menggunakan 1.000 data uji yang terdiri dari 500 berita faktual dan 500 berita hoax, dengan sumber berita dari cnnindonesia.com, kompas.com, detik.com, antaranews.com, sindonews.com, tempo.co, liputan6.com. Data uji hoax diperoleh dari situs turnbackhoax.id, beritajatim.com, indopop.id, suara.com, wowkeren.id, fimela.com, beautynesia.id, melintas.id, lintasterkini.com, celebrity.okezone.com. Hasil evaluasi menunjukkan bahwa model deteksi hoax yang dibangun mampu mencapai nilai F1-score sebesar 85%. Hasil akhir penelitian adalah sistem deteksi hoax teks berita berbahasa Indonesia yang dapat membantu masyarakat membedakan berita valid dan hoax, sehingga dapat meningkatkan literasi digital serta mengurangi penyebaran informasi palsu di ranah politik.

Kata Kunci: *Literasi Digital, Deteksi Hoax, CRISP-DM, Teks Berita Politik, F1-score*



ABSTRACT

Low levels of digital literacy in Indonesian society make it difficult for people to distinguish valid information from hoaxes, particularly in political news. This study aims to develop a hoax detection system for Indonesian-language political news using the Cross-Industry Standard Process for Data Mining (CRISP-DM) methodology. The dataset was obtained through web scraping from the official website of the Cabinet Secretariat of the Republic of Indonesia (setkab.go.id). The evaluation was carried out using 1,000 test data comprising 500 factual news articles and 500 hoax articles collected from cnnindonesia.com, kompas.com, detik.com, antaranews.com, sindonews.com, tempo.co, and liputan6.com. Hoax test data were sourced from turnbackhoax.id, beritajatim.com, indopop.id, suara.com, wowkeren.id, fimela.com, beautynesia.id, melintas.id, lintasterkini.com, and celebrity.okezone.com. The evaluation results indicate that the developed hoax detection model achieved an F1-score of 85%. The final outcome of this research is an Indonesian-language news hoax detection system that can help the public distinguish valid information from hoaxes, thereby improving digital literacy and reducing the spread of false information in the political sphere.

Keywords: *Digital Literacy, Hoax Detection, CRISP-DM, Political News Text, F1-score*

