

INTISARI

KOMBINASI FITUR *LURE* DAN *SIMILARITY* UNTUK DETEKSI *CLICKBAIT* PADA ARTIKEL BERITA BERBAHASA INDONESIA

Irbah Asfarina
20/455448/PA/19663

Salah satu strategi yang digunakan di industri media digital untuk meraih keuntungan adalah dengan merancang penulisan judul artikel agar menarik perhatian. Namun, masalah muncul ketika konten dari artikel tersebut ternyata tidak menjelaskan hal yang tertulis pada judul artikel. Tidak hanya menurunkan kepercayaan pembaca, hal itu juga membuat rasa penasaran dari pembaca menjadi tidak terpuaskan dan berkurangnya informasi yang bisa didapat oleh pembaca. Artikel semacam itu disebut sebagai *clickbait*. Penelitian yang mengembangkan metode pendeteksi *clickbait* untuk berita berbahasa Indonesia telah dilakukan, namun lebih banyak yang hanya menekankan pada fitur karakteristik dari judul berita tanpa mempertimbangkan kemiripan antara judul berita dan konten berita.

Penelitian ini menggunakan dua fitur utama untuk mendeteksi *clickbait* dalam berita berbahasa Indonesia: karakteristik judul (*lure*) dan kemiripan antara judul dan konten berita (*similarity*). Fitur *lure* diperoleh dengan menggunakan Bi-LSTM untuk menganalisis judul berita, sedangkan fitur *similarity* mengukur kemiripan antara judul dan konten menggunakan *cosine distance*. Representasi judul dan konten berita dihasilkan menggunakan Bi-LSTM dan blok Transformer Encoder. Melalui mekanisme prediksi adaptif, kedua fitur digabungkan untuk mengklasifikasi *clickbait*. Pengujian dilakukan terhadap tiga variasi fitur: *lure* saja, *similarity* saja, dan kombinasi *lure* dan *similarity*. Model terbaik didapat dengan menggunakan kombinasi kedua fitur dengan akurasi 81% dan F1 score 76,1%.

Kata kunci: Deteksi *clickbait*, Bahasa Indonesia, Bi-LSTM, blok Transformer Encoder

ABSTRACT

THE COMBINATION OF LURE AND SIMILARITY FEATURES FOR CLICKBAIT DETECTION IN BAHASA INDONESIA NEWS ARTICLES

Irbah Asfarina
20/455448/PA/19663

One strategy used in the digital media industry to generate profits is by crafting article titles to grab attention. However, a problem arises when the content of the article fails to deliver what is promised in the title. Not only does this diminish the trust of readers, but it also leaves their curiosity unsatisfied and reduces the information they can gain. Articles of this nature are referred to as clickbait. Research has been conducted to develop methods for detecting clickbait in Bahasa Indonesia news, but many focus solely on the characteristic features of headlines without considering the similarity between headlines and the content of the news.

This research utilizes two main features to detect clickbait in Bahasa Indonesia news: headline characteristics (lure) and the similarity between headlines and content (similarity). The lure feature is obtained by using Bi-LSTM to analyze news headlines, while the similarity feature measures the resemblance between headlines and content using cosine distance. Representations of headlines and news content are generated using Bi-LSTM and Transformer Encoder blocks. Through an adaptive prediction mechanism, both features are combined to classify clickbait. Testing is conducted on three variations of features: lure only, similarity only, and a combination of lure and similarity. The best model is achieved by using a combination of both features with an accuracy of 81% and an F1 score of 76.1%.

Keywords: Clickbait Detection, Bahasa Indonesia, Bi-LSTM, Transformer Encoder blocks