

## INTISARI

### Deteksi Clickbait Menggunakan Augmentasi Data dan DistilBERT

oleh

Yovita Zahra Khoerinisa

(20/462194/PA/20166)

Perkembangan teknologi telah mengubah cara penyebaran informasi, dari yang terbatas dan terkontrol menjadi bebas dan cepat di era digital. Hal ini memungkinkan penyebaran informasi yang menyesatkan (*misleading content*), yang dapat berbahaya bagi masyarakat. *Headline* berita *online* sering digunakan untuk menarik perhatian pembaca dengan *clickbait*, yaitu judul yang menyesatkan atau tidak sesuai dengan isi berita. Untuk menangani penyebaran *clickbait*, diperlukan sistem pendeteksi otomatis yang dapat memeriksa kesesuaian antara *headline* dan isi berita.

Fokus penelitian ini adalah untuk membandingkan kinerja model DistilBERT, varian BERT yang lebih ringan dan cepat. BERT memiliki kelemahan komputasi berat dan waktu pelatihan lama, sehingga DistilBERT dipilih sebagai alternatif. Penelitian ini menggunakan dataset FNC-1 yang tidak seimbang. Untuk mengatasi ketidakseimbangan ini dan mencegah bias pada model, digunakan teknik augmentasi data. Tujuan penelitian ini adalah membandingkan performa arsitektur DistilBERT dan BERT serta menilai pengaruh augmentasi data pada dataset yang tidak seimbang.

Hasil evaluasi dan perbandingan menunjukkan bahwa augmentasi data tidak berdampak signifikan pada peningkatan performa model. Sebaliknya, augmentasi dapat menyebabkan penambahan *noise* pada dataset yang malah mengurangi performa model. Selain itu, penggunaan model DistilBERT menunjukkan kemampuan untuk bersaing dengan BERT. Dengan waktu komputasi yang lebih singkat, hasil penelitian menunjukkan bahwa perbedaan performa antara BERT dan DistilBERT tidak begitu signifikan.

Kata Kunci : *Clickbait*, Deteksi *Stance*, BERT, DistilBERT, Augmentasi Data, FNC-1.

## ABSTRACT

### Clickbait Detection Using Data Augmentation and DistilBERT

by

Yovita Zahra Khoerinisa

(20/462194/PA/20166)

Technological advancements have transformed the dissemination of information from being limited and controlled to being free and fast in the digital era. This allows the spread of misleading content, which can be harmful to society. Online news headlines are often used to attract readers' attention with clickbait, misleading or mismatched headlines with the news content. To address the spread of clickbait, an automatic detection system is needed to check the consistency between headlines and news content.

This research focuses on comparing the performance of the DistilBERT model, a lighter and faster variant of BERT. BERT has the drawback of heavy computation and long training times, so DistilBERT was chosen as an alternative. This study uses the imbalanced FNC-1 dataset. To address this imbalance and prevent model bias, data augmentation techniques are used. The goal of this research is to compare the performance of the DistilBERT and BERT architectures and to assess the impact of data augmentation on the imbalanced dataset.

The evaluation and comparison results show that data augmentation does not significantly improve model performance. Instead, augmentation can add noise to the dataset, which reduces model performance. Additionally, the use of the DistilBERT model shows competitive capabilities with BERT. With shorter computation times, the study shows that the performance difference between BERT and DistilBERT is not very significant.

Keywords: Clickbait, Stance Detection, BERT, DistilBERT, Data Augmentation, FNC-1.