

INTISARI

DETEKSI *CLICKBAIT* PADA JUDUL BERITA BERBAHASA INDONESIA DENGAN METODE *FEW-SHOT LEARNING* MENGUNAKAN SETFIT

Oleh

M. Farhan Dwi Rizqi
20/459271/PA/19932

Beberapa penelitian terkait pendeteksian *clickbait* pada judul berita berbahasa Indonesia telah dilakukan menggunakan *fine-tuning* pada *pre-trained language model* sebelumnya, penelitian ini akan melakukan pendeteksian menggunakan metode *few-shot learning* karena keunggulannya dalam melakukan inferensi dengan menggunakan *data training* terbatas. Metode *few-shot learning* yang akan digunakan pada penelitian ini adalah SetFit karena dapat menghasilkan performa yang baik dengan komputasi yang cukup kecil bila dibanding dengan metode *few-shot learning* lainnya. Metode ini akan dibandingkan dengan *fine-tuning* pada IndoBERT yang merupakan *pre-trained language model* yang di-*pre-train* menggunakan bahasa Indonesia. Model dilatih dan diuji menggunakan *dataset CLICK-ID*. Akan dibandingkan hasil performa pada model SetFit menggunakan sampel data *training* 8, 32, 64, dan 100 dengan *fine-tuning* pada IndoBERT yang dilatih pada data *training* utuh dan sampel 100. Hasil penelitian menunjukkan bahwa model SetFit mendapatkan hasil yang bersaing dengan model IndoBERT yang dilatih pada data *training* utuh dan jauh lebih baik dari model IndoBERT yang dilatih dengan 100 sampel data *training* per-kelas. Dengan adanya penelitian ini, diharapkan dapat ditemukan alternatif untuk melakukan pembelajaran mesin berbasis teks, di mana bisa menggunakan metode *few-shot learning*.

Kata-kata kunci : *Few-shot Learning*, SetFit, *pre-trained language model*, *clickbait*, *CLICK-ID*, IndoBERT, *fine-tuning*.

ABSTRACT

Clickbait Detection in Indonesian News Headlines Using the Few-Shot Learning Method with SetFit

By

M. Farhan Dwi Rizqi
20/459271/PA/19932

Several studies related to clickbait detection in Indonesian news headlines have been conducted using fine-tuning on previously pre-trained language models. This research will perform detection using the few-shot learning method due to its advantages in making inferences with limited training data. The few-shot learning method used in this study is SetFit because it can achieve good performance with relatively low computational cost compared to other few-shot learning methods. This method will be compared with fine-tuning on IndoBERT, which is a pre-trained language model that has been pre-trained using the Indonesian language. The models are trained and tested using the CLICK-ID dataset. The performance of the SetFit model will be compared using training data samples of 8, 32, 64, and 100 with fine-tuning on IndoBERT trained on the full training data and a sample of 100. The results of the study show that the SetFit model achieves competitive results with the IndoBERT model trained on the full training data and performs significantly better than the IndoBERT model trained with 100 training samples per class. This research is expected to provide an alternative for text-based machine learning that utilizes the few-shot learning method.

Keywords : Few-shot Learning, SetFit, pre-trained language model, clickbait, CLICK-ID, IndoBERT, fine-tuning.