INTISARI

Penyeimbangan Data Dengan *Easy Data Augmentation* (EDA) pada Deteksi Ujaran Kebencian

Oleh

Safira Isma Aulia
21/481380/PA/20959

Dalam era digital yang terus berkembang, media sosial telah menjadi salah satu platform utama interaksi masyarakat. Namun, kemajuan ini juga disertai peningkatan ujaran kebencian yang mengancam harmoni sosial. Salah satu tantangan dalam pengembangan sistem deteksi ujaran kebencian adalah ketidakseimbangan data, di mana jumlah data ujaran kebencian jauh lebih sedikit dibanding data non-kebencian. Ketidakseimbangan ini sering kali menyebabkan bias dalam model klasifikasi dan menurunkan performa deteksi khususnya pada kelas minoritas (ujaran kebencian).

Penelitian ini bertujuan untuk meningkatkan performa model deteksi ujaran kebencian dengan mengatasi ketidakseimbangan data melalui penerapan teknik augmentasi data *Easy Data Augmentation* (EDA). Dataset yang digunakan adalah IndoToxic2024, yang berisi teks berbahasa Indonesia dari berbagai platform media sosial. Teknik EDA diterapkan menggunakan empat operasi utama, yakni *Synonym Replacement, Random Insertion, Random Swap,* dan *Random Deletion*, untuk memperluas variasi data pada kelas minoritas. Model klasifikasi dibangun dengan arsitektur *Bidirectional Long Short-Term Memory* (BiLSTM) yang memanfaatkan FastText sebagai metode ekstraksi fitur.

Hasil penelitian menunjukkan bahwa penerapan EDA berhasil secara signifikan meningkatkan performa model dalam mendeteksi kelas minoritas (ujaran kebencian), dengan peningkatan *recall* hingga 40,05% dan *F2-Score* sebesar 20,97%.

Kata Kunci: *hate speech detection, easy data augmentation, data imbalance, bidirectional long short-term memory, fasttext, text data augmentation, nlp*

ABSTRACT

Data Balancing Using Easy Data Augmentation *(EDA) in Hate Speech Detection*

By

Safira Isma Aulia
21/481380/PA/20959

In the rapidly evolving digital era, social media has become one of the main platforms for public interaction. However, this progress is accompanied by an increase in hate speech, which threatens social harmony. One of the main challenges in developing hate speech detection systems is data imbalance, where the amount of hate speech data is significantly smaller than non-hate speech data. This imbalance often leads to bias in classification models and reduces detection performance, especially for the minority class (hate speech).

This study aims to improve the performance of hate speech detection models by addressing data imbalance through the application of the Easy Data Augmentation (EDA) technique. The dataset used is IndoToxic2024, containing Indonesian text from various social media platforms. EDA is applied using four main operations: Synonym Replacement, Random Insertion, Random Swap, and Random Deletion, to expand data variation in the minority class. The classification model is built using the Bidirectional Long Short-Term Memory (BiLSTM) architecture, leveraging FastText as the feature extraction method.

The results show that applying EDA significantly improves the model's performance in detecting the minority class (hate speech), with a recall increase of up to 40.05% and an F2-Score improvement of 20.97%.

Keywords**:** *hate speech detection, easy data augmentation, data imbalance, bidirectional long short-term memory, fasttext, text data augmentation, nlp*