Abstrak

Sentiment analysis **adalah salah satu bidang dari** ***Natural Language Processing*** **(NLP) yang bertujuan mengklasifikasikan teks ke dalam sentimen positif, negatif, atau netral. Beberapa kendala yang menghalangi model untuk bekerja optimal, seperti dataset yang tidak seimbang, ketika suatu dataset memiliki beberapa kelas dengan sampel yang lebih sedikit. Untuk mengatasinya, beberapa metode melibatkan** ***oversampling*** **dan** ***data augmentation*** **untuk menambah data kelas minoritas. Selain itu,** ***cost-sensitive learning*** **adalah di mana model menangani bobot setiap kelas tanpa menambahkan data dengan perubahan dari** ***loss function*** **Cross-Entropy menjadi Focal Loss dan** ***optimizer*** **Adam menjadi NAdam. Kendala lainnya adalah dengan model berbasis BERT terbaru yang ditemukan adalah model yang berukuran besar. Model-model ini membutuhkan waktu yang lebih lama untuk diproses dan oleh karenanya membutuhkan daya komputasi yang sangat besar. Pada penelitian ini, metode-metode tersebut dievaluasi pada dataset Indonesian General** ***Sentiment analysis*** **yang tidak seimbang dan mengandung kalimat informal untuk mengukur performa BERT, IndoBERT, DistilBERT, DistilBERT-Indo, dan RoBERTa dalam menangani** ***imbalanced dataset*** **dan waktu yang dibutuhkan. Hasilnya, random oversampling mengungguli** ***data augmentation*****, dan IndoBERT mengungguli model-model lainnya dibandingkan dengan 88,46% Macro F1-Score dan 94,41% pada set data yang ditingkatkan, dengan Focal Loss dan NAdam rata-rata lebih baik dari Cross-Entropy dan Adam.**

Kata kunci : ***sentiment analysis*****, transformers,** ***imbalanced dataset*****,** ***oversampling*****,** ***cost-sensitive learning***