

## ABSTRACT

*Sentiment Analysis of Reviews Using Variants of BERT Models with Data Imbalance Handling Through Augmentation Techniques*

By

Arridho Ramadhan Firdaus

23/526962/PPA/06668

*This study evaluates the performance of three BERT-based language models: BERT, DistilBERT, and ALBERT, for sentiment analysis on Coursera user reviews, with a focus on addressing data imbalance using augmentation techniques such as Back Translation, Contextual Word Embedding, and Synonym Replacement. These techniques increase data variation, particularly for minority classes, which helps improve accuracy in sentiment classification. Results indicate that DistilBERT achieved the best performance using the synonym replacement technique, with an F1-score of 0.9869. Although DistilBERT provided the highest score for synonym replacement, models like BERT consistently delivered reliable results. Similarly, ALBERT exhibited consistent performance. Among the three BERT-based models, BERT is deemed the most suitable for sentiment analysis due to its consistently reliable outcomes.*

**Keywords:** *Sentiment Analysis, BERT, DistilBERT, ALBERT, Data Imbalance, Augmentation Techniques*

## INTISARI

### **Analisis Sentimen Ulasan Menggunakan Variasi Model BERT dengan Penanganan Ketidakseimbangan Data melalui Teknik Augmentasi**

Oleh

Arridho Ramadhan Firdaus

23/526962/PPA/06668

Penelitian ini mengevaluasi performa tiga model bahasa berbasis BERT, yaitu BERT, DistilBERT, dan ALBERT, untuk analisis sentimen pada ulasan pengguna Coursera, dengan fokus menangani ketidakseimbangan data menggunakan teknik augmentasi seperti Back Translation, Contextual Word Embedding, dan Synonym Replacement. Dengan teknik ini, variasi data meningkat terutama untuk kelas minoritas, yang membantu meningkatkan akurasi dalam klasifikasi sentimen. Hasil menunjukkan bahwa model DistilBERT memberikan performa terbaik pada teknik synonym replacement dengan hasil f1-score 0.9869, meskipun DistilBERT menghasilkan hasil tertinggi pada teknik synonym replacement, model seperti BERT memberikan hasil yang selalu konsisten begitupun dengan ALBERT hasil yang diberikan cukup konsisten. Berdasarkan ketiga model BERT tersebut model yang cocok untuk analisis sentiment yaitu BERT karena hasil yang didapatkan selalu konsisten.

**Kata Kunci:** Analisis Sentimen, BERT, DistilBERT, ALBERT, Ketidakseimbangan Data, Teknik Augmentasi