

ABSTRAK

Penggunaan media digital seperti YouTube semakin meningkat, termasuk podcast yang menyajikan berbagai topik. Namun, percakapan dalam podcast sering kali mengandung bahasa kasar yang dapat menimbulkan ketidaknyamanan bagi pendengar. Penelitian ini bertujuan mendeteksi bahasa kasar dalam konten podcast youtube menggunakan algoritma BERT dengan metodologi CRISP-DM. Dataset terdiri dari transkrip 13 video, dengan penanganan ketidakseimbangan data menggunakan SMOTE. Model BERT mencapai performa optimal pada pembagian data 80:20 dan 3 epoch, dengan Precision, Recall, dan F1-Score sempurna sebesar 1.00, serta BERTScore F1 sebesar 0.829. Penelitian ini juga menghasilkan prototipe aplikasi berbasis web yang secara otomatis memproses transkrip video youtube, mendeteksi bahasa kasar, dan menampilkan hasil analisis. Hasil penelitian ini berkontribusi dalam pengembangan teknologi analisis teks berbasis machine learning untuk menciptakan komunikasi digital yang lebih kondusif.

Kata kunci : *podcast youtube, Algoritma BERT, SMOTE*



ABSTRACT

The use of digital media such as YouTube is increasing, including podcasts that cover various topics. However, conversations in podcasts often contain offensive language that may cause discomfort to listeners. This study aims to detect offensive language in YouTube podcast content using the BERT algorithm with the CRISP-DM methodology. The dataset consists of transcripts from 13 videos, with data imbalance handled using SMOTE. The BERT model achieved optimal performance with an 80:20 data split and 3 epochs, obtaining perfect Precision, Recall, and F1-Score of 1.00, along with a BERTScore F1 of 0.829. This study also developed a web-based prototype that automatically processes YouTube video transcripts, detects offensive language, and presents the analysis results. The findings contribute to the development of text analysis technology based on machine learning to create a more conducive digital communication environment.

Keywords : youtube Podcast , BERT Algorithm, SMOTE

