

INTISARI

Komparasi Kombinasi *Word Embedding* dengan Algoritma Klasifikasi dalam Analisis Sentimen Komentar YouTube

Oleh

Laras Septiana Dewi

20/456583/PA/19770

Perkembangan teknologi digital telah mendorong penggunaan media sosial sebagai sarana untuk mengekspresikan opini publik. YouTube, salah satu platform berbagi video terbesar, menyediakan data komentar yang sangat kaya untuk dianalisis, terutama dalam konteks politik. Penelitian ini bertujuan untuk melakukan analisis sentimen terhadap komentar video siaran langsung Debat Calon Presiden (Capres) dan Calon Wakil Presiden (Cawapres) Pemilihan Umum Republik Indonesia Tahun 2024, serta membandingkan performa kombinasi metode *word embedding* (Word2Vec, GloVe, dan fastText) dengan algoritma klasifikasi (Support Vector Machine, Random Forest, Convolutional Neural Network, dan eXtreme Gradient Boosting). Data diperoleh melalui *web scraping* terhadap komentar dari kanal YouTube resmi KPU RI (@KPURepublikIndonesia), kemudian diproses melalui tahapan *text preprocessing*, pelabelan sentimen menggunakan *word embedding*, serta penanganan ketidakseimbangan kelas menggunakan metode SMOTE. Hasil penelitian pada studi kasus menunjukkan bahwa komentar pengguna secara umum cenderung bersentimen positif. Kombinasi fastText dengan SMOTE dan algoritma XGBoost memberikan performa terbaik dengan akurasi sebesar 93,63%, diikuti oleh Random Forest (93,34%) dan CNN (91,26%). Sementara itu, model SVM memperoleh akurasi terendah (90,39%). Berdasarkan hasil yang diperoleh, dapat disimpulkan bahwa pemilihan metode representasi teks dan algoritma klasifikasi yang tepat memiliki peran penting dalam meningkatkan akurasi analisis sentimen, khususnya pada data berbahasa Indonesia di media sosial.

Kata Kunci: Analisis Sentimen, *Word Embedding*, SMOTE, Komentar YouTube

ABSTRACT

Comparison of Word Embedding and Classification Algorithm Combinations in YouTube Comment Sentiment Analysis

By

Laras Septiana Dewi

20/456583/PA/19770

Digital technology development has encouraged the use of social media as a means to express public opinion. YouTube, one of the largest video-sharing platforms, provides a great resource of valuable comment databases for analysis, especially in the context of politics. This research aims to conduct sentiment analysis of live video comments on the Presidential Candidate and Vice Presidential Candidate Debate of the Republic of Indonesia Election in 2024 and compare the performance of a combination of word embedding methods (Word2Vec, GloVe, and fastText) with classification algorithms (Support Vector Machine, Random Forest, Convolutional Neural Network, and eXtreme Gradient Boosting). The data was obtained through web scraping of comments from the official YouTube channel of KPU RI (@KPURepublikIndonesia), then processed through the stages of text preprocessing, sentiment labeling using word embedding, and handling class imbalance using the SMOTE method. The research results in the case study show that user comments generally tend to have positive sentiments. The combination of fastText with SMOTE and XGBoost achieved the highest accuracy at 93.63%, followed by Random Forest (93.34%) and CNN (91.26%). Meanwhile, the SVM model is at the lowest position with 90.39% accuracy. According to the results obtained, it can be concluded that the selection of the right text representation method and classification algorithm has an important role in improving the accuracy of sentiment analysis, especially on Indonesian-language data on social media.

Keywords: Sentiment Analysis, Word Embedding, SMOTE, YouTube Comments