

Model Gabungan fastText-CNN-BiLSTM untuk Analisis Sentimen pada Data Opini Berbahasa Indonesia = Hybrid Model fastText-CNN-BiLSTM for Sentiment Analysis on Indonesian Opinion Data

Suci Fitriyani, author

Deskripsi Lengkap: <https://lib.ui.ac.id/detail?id=9999920531273&lokasi=lokal>

Abstrak

Analisis sentimen merupakan studi komputasi untuk menganalisis opini seseorang terhadap suatu entitas yang diekspresikan dalam sebuah teks. Tersedia cukup banyak model machine learning terutama deep learning yang dapat digunakan untuk melakukan analisis sentimen seperti Convolutional Neural Network (CNN) dan Bidirectional Long Short-Term Memory (BiLSTM). Pada dasarnya, model deep learning tidak dapat memproses langsung sebuah data dalam bentuk teks sehingga diperlukan metode untuk mentransformasi teks menjadi tensor numerik seperti word embedding. Pada penelitian ini, diajukan model gabungan CNN-BiLSTM dengan word embedding fastText untuk melakukan analisis sentimen. Model tersebut dilatih menggunakan data tweet berbahasa Indonesia tentang opini masyarakat mengenai rencana subsidi pembelian kendaraan listrik di Indonesia. Data tersebut diklasifikasikan menjadi sentimen positif, negatif, dan netral dan ditemukan bahwa komposisi dari ketiga sentimen tersebut tidaklah seimbang (imbalanced dataset) dimana kelas positif memiliki lebih sedikit data dibanding kelas kelas lainnya. Untuk mengatasi masalah tersebut, digunakan metode resampling SMOTE agar jumlah data pada kelas positif dapat mengimbangi kelas lainnya. Model fastText CNN-BiLSTM diukur performanya dengan melihat nilai akurasi, precision, recall, dan f1-score. Dari hasil penelitian didapat bahwa model gabungan CNN-BiLSTM memberikan nilai akurasi, precision, recall, dan f1-score yang paling baik dibanding model CNN dan BiLSTM saja. Model-model yang menggunakan word embedding fastText juga memberikan performa yang lebih baik dibanding model tanpa fastText (menggunakan word embedding standar). Secara keseluruhan, model gabungan fastText CNN-BiLSTM ditemukan memiliki performa yang lebih baik dibandingkan dengan model-model lainnya.

.....Sentiment analysis is a computational study to analyze person's opinion about an entity expressed in text. There are several machine learning models, especially deep learning models that can be used for sentiment analysis, such as Convolutional Neural Network (CNN) and Bidirectional Long Short-Term Memory (BiLSTM). Essentially, deep learning models cannot directly process textual data and they need a method to transform text into numerical tensors such as word embedding. In this research, a hybrid model CNN-BiLSTM with fastText word embedding is proposed for sentiment analysis. The model is trained using Indonesian tweets data regarding public opinions on the plan for subsidizing the purchase of electric vehicles in Indonesia. The data is classified into positive, negative, and neutral sentiments, and it is found that the composition of these sentiments is imbalanced, with the positive class having fewer data compared to the other classes. To address this issue, the SMOTE resampling method is used to balance the data in the positive class with the other classes. The performance of the fastText-CNN BiLSTM model is measured by accuracy, precision, recall, and f1-score. The research results show that the hybrid model CNN-BiLSTM achieves the highest accuracy, precision, recall, and f1-score compared to the single models CNN and BiLSTM. Models with fastText word embedding also outperform models without fastText (with standard word embedding). Overall, the hybrid model fastText-CNN-BiLSTM is found to outperform other models in

terms of performance.