

Abstrak

Pesatnya perkembangan teknologi dan internet telah mengakibatkan semakin banyaknya berita yang tersedia secara daring. Hal ini menimbulkan tantangan dalam mengelompokkan berita berdasarkan topiknya karena jumlahnya yang terus meningkat. Penelitian ini bertujuan untuk membangun sistem klasifikasi topik berita berbahasa Indonesia menggunakan pendekatan *FastText pretrained* dan mengimplementasikannya ke dalam aplikasi prediksi berbasis web. Data yang digunakan diperoleh dari 4.189 artikel berita yang telah melalui proses preprocessing, meliputi *case folding*, tokenisasi, penghapusan *stopword*, dan *stemming* menggunakan *library Sastrawi*. Model klasifikasi dibangun menggunakan arsitektur *deep learning* dengan *pretrained FastText embedding* berdimensi 300, *GlobalAveragePooling1D*, dan *output layer* beraktivasi *softmax*. Dataset dibagi menjadi 80% data latih (3.350 data) dan 20% data uji (838 data), dengan pelabelan manual ke dalam 13 kategori topik seperti Ekonomi, Politik, Kesehatan, dan Olahraga. Hasil evaluasi model menunjukkan akurasi sebesar 80,4%, dengan *f1-score* juga cukup tinggi pada beberapa kategori seperti Cuaca memperoleh *f1-score* tinggi (0.97), sedangkan kategori Sosial & Budaya dan Politik memiliki *f1-score* relatif rendah. Model ini diimplementasikan dalam aplikasi berbasis *Streamlit* untuk mempermudah prediksi topik terhadap berita baru secara interaktif.

Kata kunci: Berita Indonesia, *Deep Learning*, *FastText*, Klasifikasi Teks, NLP, *Pretrained Embedding*, *Streamlit*

Abstract

The rapid development of technology and the internet has resulted in an increasing number of news items available online. This poses a challenge in grouping news items based on their topics due to the ever-increasing number. This study aims to build an Indonesian-language news topic classification system using the pretrained FastText approach and implement it into a web-based prediction application. The data used were obtained from 4,189 news articles that have undergone preprocessing, including case folding, tokenization, stopword removal, and stemming using the Sastrawi library. The classification model was built using a deep learning architecture with a 300-dimensional pretrained FastText embedding, GlobalAveragePooling1D, and a softmax-activated output layer. The dataset was divided into 80% training data (3,350 data sets) and 20% test data (838 data sets), with manual labeling into 13 topic categories such as Economy, Politics, Health, and Sports. The model evaluation results showed an accuracy of 80.4%, with a relatively high f1-score in several categories, such as Weather, which achieved a high f1-score (0.97), while the Social & Culture and Politics categories had relatively low f1-scores. This model was implemented in a Streamlit-based application to facilitate interactive topic prediction for new news.

Keywords: Deep Learning, FastText, Indonesian News, NLP, Pretrained Embedding, Streamlit, Text Classification