

# Klasifikasi Dokumen dan Ekstraksi Lokasi pada Berita Bencana Alam dengan Pendekatan Neural Network dan Pre-Trained Language Model = Document Classification and Location Extraction in Natural Disaster News with Neural Network Approach and Pre-Trained Language Model

Machffud Tra Harana Vova, author

Deskripsi Lengkap: <https://lib.ui.ac.id/detail?id=9999920542453&lokasi=lokal>

---

## Abstrak

Indonesia merupakan negara yang wilayahnya sering mengalami bencana alam. Salah satu penanganan bencana alam adalah pengumpulan informasi berita bencana seperti artikel atau koran, yang mana berguna untuk meningkatkan readability. Meskipun begitu, sekadar pengumpulan artikel saja cukup sulit karena identifikasinya dapat memakan waktu serta makna yang termuat pada berita juga masih perlu diserap. Oleh karena itu perlu dilakukan klasifikasi dokumen untuk memilih teks dokumen yang relevan dengan bencana alam, kemudian dari teks dokumen yang relevan dilakukan ekstraksi informasi. Penelitian mengenai klasifikasi teks bencana alam serta ekstraksi informasi yang sudah dilakukan masih menggunakan pendekatan pembelajaran mesin tradisional serta belum memanfaatkan pre-trained model berbasis bahasa Indonesia. Penggunaan pre-trained model dan pendekatan deep learning sendiri sering memperoleh performa yang lebih baik, sehingga ada kemungkinan performa yang dihasilkan dapat ditingkatkan. Dalam penelitian ini dilakukan eksperimen menggunakan pre-trained word embedding seperti Word2Vec dan fastText, pendekatan deep learning seperti BERT dan BiLSTM untuk task klasifikasi. Hasil dengan pendekatan pembelajaran mesin tradisional dengan BoW yang sudah direproduksi menjadi yang terbaik hampir secara keseluruhan, meskipun jenis classifier yang digunakan adalah MLP yang mana sudah menerapkan deep learning karena memiliki beberapa neuron. Kemudian pada penggunaan model pre-trained seperti BERT, terdapat keterbatasan panjang masukan. Keterbatasan ini dapat ditangani dengan membuat representasi dokumen menjadi lebih pendek menggunakan metode peringkasan teks. Hasil representasi ringkasan dokumen dalam penelitian ini mampu meningkatkan performa akurasi klasifikasi baik pada pendekatan pembelajaran mesin tradisional maupun deep learning. Penelitian ini juga melakukan eksperimen penggunaan pre-trained model yang sudah fine-tuned untuk task ekstraksi lokasi seperti NER dan dependency parsing berbasis bahasa Indonesia, meskipun belum dihasilkan performa yang cukup baik.

.....Indonesia is a country whose often experiences natural disasters. One way to deal with natural disasters is to collect disaster news information such as articles or newspapers, which are useful for increasing readability. Even so, just collecting articles is quite difficult because identification can take time and the meaning contained in the news still needs to be absorbed. Therefore, it is necessary to classify documents to select document texts that are relevant to natural disasters, then extract information from the relevant document texts. Research on natural disaster text classification and information extraction that has been carried out still uses the traditional machine learning approach and has not yet utilized Indonesian language-based pre-trained models. The use of pre-trained models and deep learning approaches themselves often get better performance, so there is a possibility that the resulting performance can be improved. In this study, experiments were carried out using pre-trained word embedding such as Word2Vec and fastText, deep learning approaches such as BERT and BiLSTM for classification tasks. The results with traditional machine learning approaches with reproducible BoW are the best almost overall, even though the type of

classifier used is MLP which already implements deep learning because it has few neurons. Then in the use of pre-trained models such as BERT, there are limitations to the length of the input. This limitation can be overcome by making the document representation shorter using the text summary method. The results of the document summary representation in this study were able to improve the performance of classification accuracy in both traditional and deep learning machine learning approaches. This study also conducted experiments using pre-trained models that had been fine-tuned for location extraction tasks such as NER and Indonesian language-based dependency parsing, although they did not produce sufficiently good performance.