

ABSTRAK

Inovasi berbagai teknologi di era ini, telah membawa manusia ke peradaban yang serba mudah, terutama dalam bidang pengelolaan informasi. Peran *Internet*, sebagai media untuk melakukan distribusi berbagai informasi, telah membantu manusia dalam menyerap informasi, pengetahuan, dan edukasi secara cepat dan efisien dalam skala global. Namun perkembangan informasi yang luar biasa ini, juga menyebabkan dampak negatif dalam penyebaran berita hoaks secara berlebihan, yang tentunya menjadi masalah bagi pengguna internet.

Untuk mengatasi hal ini, banyak teknik manual yang bisa dilakukan untuk menghindari berita hoaks seperti, mengecek kebenaran dengan membandingkan dengan berita lainnya, melihat keaslian situs web asli, memeriksa keaslian foto, dan teknik lainnya. Namun cara-cara ini tentunya tidak bisa diimplementasikan kepada seluruh masyarakat, dikarenakan pengetahuan dan kemampuan pengguna internet berbeda antara satu dengan lainnya, maka dari itu banyak penelitian yang dilakukan untuk mencari cara deteksi berita secara otomatis dengan menggunakan bantuan dari *Artificial Intelligence* dan *Machine Learning*.

Dalam penelitian ini, penulis mengusulkan untuk mendeteksi hoaks pada *dataset* “Indonesia False News(Hoax)” digabungkan dengan *dataset* yang telah diambil melalui metode *scraping*. Kemudian, model pre-trained *Bidirectional Encoder Representations from Transformers*(BERT) yaitu model BERT-*base-small* dan model DistilBERT digunakan untuk embedding. *Convolutional Neural Network*(CNN) digunakan untuk melakukan klasifikasi berita hoaks dengan menggunakan *dataset* bernama “Indonesia False News(Hoax)” dan digabungkan dengan *dataset* yang diperoleh melalui web *Scraping*. Model BERT – CNN dengan model pre-trained DistilBERT memiliki *hyperparameter batch size* 128, *learning rate* 0.00001, 5 *epoch* dan *Layer Convolution Neural Network* sebanyak 18-Layers berhasil mendapatkan akurasi yang optimal sebesar 89%.

Kata Kunci: BERT, *Convolutional Neural Network*, *Text Classification*, *Hoax*, *Scraping*, *Artificial Intelligence*, *Machine Learning*.

ABSTRACT

The development of numerous technologies in this age has aided human civilization, particularly in the area of information systems. Because of the Internet's role as a way to distribute information, humans have been able to quickly and effectively integrate information, knowledge, and education around the world.

To overcome this, various manual procedures can be used, such as confirming the validity of the news by comparing it to other news, checking the legitimacy of the originating website, verifying the authenticity of images, and some other methods. Because internet users have different levels of knowledge and skills, these methods can't be applied to the entire public, therefore numerous experiments have been conducted to develop ways to employ artificial intelligence and machine learning to automatically detect news.

In this study, the authors propose to detect hoaxes on the "Indonesia False News (Hoax)" dataset combined with the dataset that has been retrieved through the scraping method. The embedding process is then performed out using pre-trained BERT models, specifically the BERT-base-small model and DistilBERT model. Using a dataset called "Indonesia False News (Hoax)" and a dataset obtained by web scraping, Convolutional Neural Network (CNN) is used to classify hoax news. The BERT - CNN model with the pre-trained DistilBERT model has a hyperparameter batch size of 128, a learning rate of 0.00001, 5 *epochs* and an 18-Layers Convolution Neural Network Layer managed to get optimal accuracy of 89%.

Keyword: BERT, Convolutional Neural Network, Classification, Hoax, Scraping, Artificial Intelligence, Machine Learning.