

## ABSTRAK

Penyebaran berita *hoax* saat ini sedang marak terjadi, bahkan sudah masuk fase yang cukup mengkhawatirkan. Hal ini berdampak buruk bagi masyarakat seperti pengaburan fakta, hilangnya kepercayaan masyarakat dan dapat mengakibatkan perpecahan antar masyarakat. Banyaknya berita *hoax* yang tersebar membuat sulitnya masyarakat dalam memilah dan membedakan mana berita *hoax* atau bukan. Ada beberapa masyarakat yang kritis terhadap berita yang tersebar, lalu melakukan pengecekan apakah berita tersebut *hoax* atau bukan. Namun, tidak sedikit pula masyarakat yang langsung percaya terhadap berita yang disampaikan. Maka dari itu, diperlukan proses pengklasifikasian suatu konten berita apakah berita tersebut termasuk berita *hoax* atau fakta.

Penelitian ini menggunakan algoritma *Convolutional Neural Network* (CNN) dengan pembobotan *word2vec*. CNN adalah salah satu algoritma dari metode *deep learning*. Algoritma ini merupakan pengembangan dari *Multi Layer Perceptron* yang didesain untuk mengolah data dua dimensi. Namun, pada penelitian ini menggunakan algoritma CNN untuk klasifikasi berita *hoax* dengan bentuk data teks atau satu dimensi. Proses yang dilakukan yaitu melakukan pengumpulan data menggunakan metode *web scraping* dari situs *turnbackhoax.id* dan dilanjutkan dengan proses labeling data menjadi 2 kelas (*hoax* dan fakta). Lalu, data dilakukan proses *text processing*, pembobotan kata (*word embedding*) menggunakan metode *word2vec* serta melakukan proses *training* dan *testing* untuk membuat model klasifikasi CNN. Model yang dihasilkan akan dilakukan pengujian menggunakan tabel *confusion matrix* dan kurva ROC serta menggunakan metode validasi *k-fold cross validation* dengan nilai  $k=5$ . Hasil pengujian tersebut akan menghasilkan nilai rata-rata dari akurasi, presisi, recall dan AUC.

Berdasarkan hasil pengujian menggunakan *confusion matrix* menghasilkan nilai rata-rata akurasi sebesar 86,56%, presisi sebesar 86,35% dan recall sebesar 86,36%. Sedangkan untuk mengujian menggunakan kurva ROC menghasilkan rata-rata nilai AUC sebesar 0.864 dari nilai FPR dan TPR secara urut sebesar 0.156 dan 0.883. Nilai AUC yang tersebut berada pada rentang nilai 0.80 - 0.90 yang menunjukkan bahwa model CNN yang dihasilkan termasuk ke dalam kategori *good classification*.

**Kata kunci :** Berita *Hoax*, Klasifikasi Teks, *Deep Learning*, *Word2vec*, *Convolutional Neural Network*, *Confusion Matrix*, Kurva ROC, *K-Fold Cross Validation*