

ABSTRAK

KLASIFIKASI TEKS PADA ARTIKEL BERITA MENGGUNAKAN METODE *LONG SHORT TERM MEMORY* (LSTM)

Iwan La Udin¹, Firman Tempola², Abdul Mubarak³

^{1,2,3}Program Studi Teknik Informatika, Fakultas Teknik, Universitas Khairun
Jl.Jati Metro, Kota Ternate Selatan

Email: ¹iwanlaudin01@gmail.com, ²firman.tempola@unkhair.ac.id,
³amuba029@unkhair.ac.id

Berita yang disalurkan lewat sistem berbasis internet atau portal berita umumnya diklasifikasikan berdasarkan kategori tertentu, seperti politik, olahraga, ekonomi, hiburan, teknologi, kesehatan dan lain-lain. Saat ini kategorisasi berita dilakukan secara manual, sehingga kategorisasi berita dilakukan dengan membaca keseluruhan isi berita. Untuk mengatasi masalah tersebut dibutuhkan sistem klasifikasi artikel berita Bahasa Indonesia secara otomatis berdasarkan kategori yang telah ditentukan. Penelitian ini menggunakan pendekatan *Natural Language Processing* (NLP) dan menerapkan arsitektur *Long Short Term Memory* (LSTM). Penelitian ini dilakukan dengan beberapa skenario pengujian diantaranya (1) *tuning hyperparameter learning rate* 0,01 dan 0,001, (2) *stemming* dan tanpa *stemming*, (3) serta rasio perbandingan *dataset* 60:40, 70:30, 80:20, dan 90:10. Pengujian menggunakan 10.000 *dataset* dengan 5 kategori, diukur menggunakan *matrix accuracy*, *precision*, *recall* dan *f-measure*. Dari tiga skenario menghasilkan tujuh model pelatihan. Model kedua dengan *tuning hyperparameter learning rate* 0,001 tanpa menerapkan *stemming*, rasio *dataset* 90:10 menghasilkan akurasi tertinggi 90,7% dengan rata-rata nilai *precision*, *recall* dan *f-measure* adalah 91%. Model ketiga dan keempat dengan menerapkan *stemming*, tidak menunjukkan peningkatan performa dengan tingkat akurasi keduanya sebesar 89%. Pada model kelima dengan perbandingan *dataset* 60:40 menghasilkan akurasi 90% sedangkan model keenam dan model ketujuh dengan perbandingan 70:30 akurasinya sebesar 79% dan 80:20 akurasinya sebesar 88%.

Kata kunci: *klasifikasi teks, long short-term memory, artikel berita*

ABSTRACT

CLASSIFICATION OF TEXT ON NEWS ARTICLES USING LONG SHORT TERM MEMORY (LSTM) METHOD

Iwan La Udin¹, Firman Tempola², Abdul Mubarak³

^{1,2,3}Informatics Engineering Study Program, Faculty of Engineering Khairun University
Jl. Jati Metro, Kota Ternate Selatan

Email: ¹iwanlaudin01@gmail.com, ²firman.tempola@unkhair.ac.id,
³amuba029@unkhair.ac.id

News Articles. News that is distributed through internet-based systems or news portals is generally classification based on certain categories, such as politics, sports, economy, entertainment, technology, health and others. Currently news categorization is done manually, so news categorization is done by reading the whole news content. To overcome this problem, an automatic Indonesian news article classification system is needed based on a predetermined category. This study uses a Natural Language Processing (NLP) approach and applies the Long Short Term Memory (LSTM) architecture. This research was conducted with several test scenarios including (1) tuning hyperparameter learning rate 0.01 and 0.001, (2) stemming and without stemming comparison ratios of dataset 60:40, 70:30, 80:20, and 90:10. The test uses 10,000 datasets with 5 categories, measured using matrix accuracy, precision, recall and f-measure. Of the three scenarios resulted in seven training models. The second model with a tuning hyperparameter learning rate 0.001 without applying stemming, ratio of dataset 90:10 produces the highest accuracy of 90.7% with an average precision, recall and f-measure 91%. The third and fourth models by applying stemming did not show an increase in performance with an accuracy rate of 89% for both. The fifth model with a dataset 60:40 produces 90% accuracy, while the sixth model and seventh model with a ratio of 70:30 have 79% accuracy and 80:20 accuracy is 88%.

Keywords: Text Classification, Long Short-Term Memory,