

ABSTRAK

Dalam karya sastra berbentuk prosa, terdapat suatu unsur intrinsik yaitu latar suasana. Latar suasana yang dibuat penulis diharapkan dapat diterima pembaca/penikmat cerita agar lebih memahami cerita dan juga merasakan karakter-karakter yang ada pada cerita. Identifikasi suasana pada cerita bukan hanya diaplikasikan demi kepentingan pembaca/penikmat cerita, namun juga untuk kebutuhan transformasi dari bentuk karya sastra prosa tertulis ke bentuk lainnya seperti pementasan teater, iklan, musikalisasi, dan lainnya. Terdapat 630 data potongan cerita untuk penelitian ini yang didapatkan dari *web-scraping* di berbagai situs kumpulan cerita pendek bahasa Indonesia yang tidak sensitif terhadap hak cipta seperti cerpenmu.com, yang kemudian diberikan anotasi suasana oleh beberapa anotator yang merupakan pembaca aktif cerita pendek. Klasifikasi suasana didasarkan oleh nilai vektor-vektor tiap kata dalam susunan potongan cerita, atau dapat disebut *Word Embedding*. Model bahasa *Word Embedding* yang diaplikasikan yaitu dengan model *pre-trained* fastText versi wiki Indonesia yang dilanjutkan training dengan *domain-specific corpus*, yaitu dari cerita pendek pada dataset train dan test. Metode klasifikasi teks menggunakan *Long Short-Term Memory* dan *Bidirectional Long Short-Term Memory*. Pengujian menggunakan 630 data dengan model LSTM dan Bi-LSTM yang difokuskan pada variasi arsitektur layer-layernya, dan tuning parameter. Didapatkan performa terbaik yaitu dengan akurasi sebesar 60.8% untuk dataset cerita dengan 4 kelas yang sudah dilakukan *upsample* menggunakan arsitektur Bi-LSTM yang menggunakan regularisasi L2.

ABSTRACT

In a literary work in the form of prose, there is an intrinsic element, namely the setting of the atmosphere. The setting of the atmosphere created by the author is expected to be accepted by the reader/connoisseur of the story in order to better understand the story and also feel the characters in the story. The identification of the atmosphere in the story is not only applied for the benefit of the readers/connoisseurs of the story, but also for the need for transformation from the form of written prose literature to other forms such as theater performances, advertisements, musicals, and others. There are 630 story snippets for this research which were obtained from web-scraping on various Indonesian short story collection sites that are not sensitive to copyright such as cerpenmu.com, which were then annotated by several annotators who are active readers of short stories. The atmosphere classification is based on the value of the vectors of each word in the arrangement of story pieces, or it can be called Word Embedding. The Word Embedding language model that is applied is the fastText version of the Indonesian wiki pre-trained model which is followed by training with a domain-specific corpus, namely from short stories on the train and test datasets. The text classification method uses Long Short-Term Memory and Bidirectional Long Short-Term Memory. The test uses 630 data with LSTM and Bi-LSTM models which are focused on variations in the architecture of the layers, and tuning parameters. The best performance was obtained with an accuracy of 60.8% for the story dataset with 4 classes that had been upsampled using the Bi-LSTM architecture using L2 regularization.