

ABSTRAK

PERBANDINGAN NORMALISASI KATA SLANG BAHASA INDONESIA MENGUNAKAN MODEL *FASTTEXT* & *WORD2VEC* DENGAN PENDEKATAN *NATURAL LANGUAGE PROCESSING*

Rifqah Nur Surayya M.Jen¹, Muhammad Fhadli², Syarifuddin N. Kapita³
Program Studi Informatika, Fakultas Teknik, Universitas Khairun
Jl. Jati Metro, Kota Ternate
E-mail: skymystery03@gmail.com¹, muhammadfhadli@unkhair.ac.id²,
syarifuddininkapita@unkhair.ac.id³

Penggunaan kata-kata gaul sering kali digunakan sebagai sarana komunikasi di media sosial seperti *Twitter*, namun menjadi masalah bagi kalangan tertentu karena sulit dimengerti jika diucapkan di luar konteks. Hal ini dapat menyebabkan komunikasi menjadi kurang efektif, terutama bagi yang tidak terbiasa dengan *slang* tersebut. Oleh karena itu, pendekatan normalisasi kata diperlukan untuk menerjemahkan kata ke dalam bahasa formal agar lebih dipahami masyarakat. *Natural Language Processing* (NLP) adalah teknik komputasi yang menganalisis dan merepresentasikan teks atau bahasa lisan untuk mencapai pemrosesan mirip manusia. Penelitian ini fokus pada teknik ekstraksi fitur seperti *FastText* dan *Word2Vec* untuk memetakan kata ke vektor numerik. Hasil pengujian kata *slang* menunjukkan *FastText* memiliki kemiripan tertinggi 0.9934859978 dan terendah 0.8928895496, sementara *Word2Vec* memiliki kemiripan tertinggi 0.9977979123 dan terendah 0.0975351095. Waktu yang dibutuhkan *FastText* untuk *training* adalah 0.432 detik dan untuk normalisasi 0.016 detik, sedangkan *Word2Vec* memerlukan 0.027 detik untuk *training* dan 0.006 detik untuk normalisasi.

Kata kunci: *Kata Slang, Natural Language Processing, Word2Vec, FastText*

ABSTRAC

COMPARISON OF INDONESIAN SLANG WORD NORMALIZATION USING *FASTTEXT* & *WORD2VEC* MODELS WITH *NATURAL LANGUAGE PROCESSING* APPROACH

The use of slang words is often used as a means of communication on social media such as Twitter, but it becomes a problem for certain circles because it is difficult to understand if spoken out of context. This can cause communication to be less effective, especially for those who are not familiar with the slang. Therefore, a word normalization approach is needed to translate the words into formal language for better understanding. Natural Language Processing (NLP) is a computational technique that analyzes and represents text or spoken language to achieve human-like processing. This research focuses on feature extraction techniques such as FastText and Word2Vec to map words to numerical vectors. The slang word testing results show FastText has the highest similarity of 0.9934859978 and the lowest of 0.8928895496, while Word2Vec has the highest similarity of 0.9977979123 and the lowest of 0.0975351095. The time required by FastText for training is 0.432 seconds and for normalization is 0.016 seconds, while Word2Vec requires 0.027 seconds for training and 0.006 seconds for normalization.

Keywords: *Slang Words, Natural Language Processing, Word2Vec, FastTex*