

**DETEKSI KALIMAT TIDAK BAKU BERDASARKAN KAMUS BESAR
BAHASA INDONESIA DENGAN METODE *COSINE SIMILARITY*
(Studi Kasus *Text Short Message Service*)**

¹*Lukie Perdanasari (11 1065 1187),*
² *Deni Arifianto, S.Kom,* ³ *Bagus Setya R, S.T, M.Kom*
Program Studi Teknik Informatika Fakultas Teknik
Universitas Muhammadiyah Jember
E-mail : lukieperdanasari.kuliah@gmail.com

ABSTRAK

Short Message Service atau SMS adalah salah satu media komunikasi berupa pesan singkat yang sering digunakan masyarakat saat ini, bahkan antar mahasiswa dengan dosen SMS merupakan media yang masih dianggap sopan untuk berkomunikasi. Namun, banyaknya gaya penulisan dan gaya bahasa yang digunakan dalam mengirim pesan sehingga dapat mempengaruhi kalimat SMS yang digunakan menjadi kalimat tidak baku. Penelitian ini membahas kemiripan data SMS baru dengan SMS yang telah dinyatakan baku dan tidak baku untuk mendeteksi kalimat tidak baku. Acuan untuk menentukan kalimat tidak baku adalah posisi nilai tertinggi dari hasil perhitungan *cosine similarity*. Hasil pengujian dari 30 data *testing* dari dokumen SMS menunjukkan bahwa 27 bernilai benar dan 3 salah, sehingga tingkat *error* sebesar 10%. *Cosine similarity* mampu mengklasifikasi kalimat baku dan tidak baku pada teks SMS dengan tingkat *accuracy* 90%, *precision* 100%, dan *recall* 83%.

Kata kunci : baku, *Short Message Service*, *cosine similarity*, *accuracy*, *precision*, *recall*

**DETECTING NONSTANDARD SENTENCE BASED ON INDONESIAN
DICTIONARY BY USING COSINE SIMILARITY
(A Case Study *Text Short Message Service*)**

¹Lukie Perdanasari (11 1065 1187),
² Deni Arifianto, S.Kom, ³ Bagus Setya R, S.T, M.Kom
Program Studi Teknik Informatika Fakultas Teknik
Universitas Muhammadiyah Jember
E-mail : lukieperdanasari.kuliah@gmail.com

ABSTRACT

Short Message Service or SMS is one of communication media in a form of short message that people often use nowadays. Among students and lecturers, even, texting is still considered polite to communicate each other. However, due to a number of writing and language styles used in texting, it affects the sentences to the nonstandard one. This study discusses the similarity between new SMS data and those determined as standard and nonstandard ones to detect nonstandard sentences. The criterion to decide nonstandard sentences is the highest position of the result of Cosine Similarity count. The result of the test is among 30 testing data from the document of SMS shows that 27 were correct while another 3 were incorrect, and the error level was 10%, therefore. Cosine Similarity can classify standard and nonstandard sentences of texting with the level of accuracy 10%, precision 100%, and recall 83%.

Keywords : *standard, Short Message Service, cosine similarity, accuracy, precision, recall*