

## ABSTRAK

# PENERAPAN NORMALISASI TEKS UNTUK KLASIFIKASI SENTIMEN TEKS BAHASA INDONESIA DENGAN *MACHINE LEARNING*

Oleh

Indah Purwanti

20110007

Analisis sentimen adalah fokus utama dalam pengolahan bahasa alami (NLP) yang bertujuan mengenali perasaan atau sentimen berdasarkan opini dalam teks. Dataset yang berasal dari media *online* sering mengandung kata *slang* dan kesalahan penulisan yang menjadi tantangan dalam klasifikasi teks sentimen karena bentuknya yang tidak standar dan mengandung banyak *noise*. *Noise* pada data ini dapat mengurangi akurasi model analisis sentimen karena model kesulitan mengenali pola-pola teks. Oleh karena itu, normalisasi teks menjadi langkah penting dalam *preprocessing* untuk mengurangi *noise*, memastikan konsistensi data, dan meningkatkan performa model. Penelitian ini bertujuan untuk menganalisis pengaruh penerapan normalisasi teks pada tahap *preprocessing* terhadap akurasi model sentimen analisis bahasa Indonesia, dengan membuat korpus *slang* dari kombinasi kamus internet dan kata *slang* pada dataset. Data yang digunakan adalah *review* produk dari Lazada dan Tokopedia di platform *Kaggle*. Hasil penelitian menunjukkan bahwa penerapan normalisasi teks meningkatkan akurasi model sentimen. Model *Naïve Bayes* memiliki akurasi yang lebih baik daripada model *Decision Tree*, dengan akurasi model *Naïve Bayes* setelah dilakukan normalisasi dapat meningkatkan akurasi dari 88,03 % menjadi 88,79% pada data Lazada dan dari 88,23 % menjadi 90,55% pada data Tokopedia. Namun, model *Decision Tree* pada data Tokopedia mengalami penurunan kecil dari 88,32 % menjadi 88,12%. Penerapan korpus *slangword* juga meningkatkan *recall* untuk kasus negatif, meskipun *precision* cenderung menurun setelah normalisasi, menunjukkan *trade-off*. Meskipun demikian, model tetap efektif dalam mengidentifikasi sentimen positif setelah normalisasi teks.

Kata kunci: *Klasifikasi Sentimen, Machine Learning, Normalisasi teks, Teks Bahasa Indonesia*