

ABSTRAK

Lirik lagu merupakan salah satu elemen yang sangat penting dalam menggambarkan ekspresi dan emosi yang ada pada lagu tersebut. Proses klasifikasi lagu dapat dilakukan salah satunya dengan menggunakan lirik lagu sebagai media untuk mengklasifikasikan emosi yang terkandung di dalamnya. Tingkat akurasi proses klasifikasi lagu dapat dipengaruhi oleh proses preprocessing yang terdiri dari case folding, stop word filtering, tokenizing, dan lemmatization. Proses lemmatization sendiri merupakan salah satu proses penting dalam tahapan preprocessing dimana sebuah kata akan dikembalikan kedalam bentuk kamusnya sehingga dapat mereduksi dan memperjelas fitur untuk melakukan proses klasifikasi, sehingga proses klasifikasi dapat dilakukan lebih cepat. Dalam penelitian sebelumnya terkait Bahasa Jawa belum ada lemmatizer khusus untuk Bahasa Jawa sehingga pada penelitian ini lemmatizer dibangun sendiri menggunakan metode Dictionary Lookup. Penggunaan pembobotan TF-IDF dan Support Vector Machine (SVM) dilakukan pada penelitian ini dengan pertimbangan, dapat membandingkan kinerja sistem antara data yang melalui tahap lemmatization dan tidak. Proses klasifikasi diawali dengan melakukan pelabelan manual, kemudian data melalui tahap preprocessing, ekstraksi fitur, pembobotan, training dan klasifikasi menggunakan SVM. Pada proses klasifikasi emosi lagu berbahasa Jawa menggunakan SVM, lemmatization, dan TF-IDF yang dilakukan dapat mereduksi jumlah fitur sebanyak 5.89% dimana data yang melalui tahap lemmatization memiliki akurasi yang lebih baik 0.35% dari data yang tidak melalui tahap lemmatization.

Kata Kunci: Bahasa Jawa, *Lemmatizer Jawa*, *Preprocessing*, Klasifikasi Teks