

KLASIFIKASI BAHASA YANG MIRIP (BAHASA INDONESIA DAN BAHASA MALAYSIA) MENGGUNAKAN METODE *SUPPORT VECTOR MACHINE*

RIRIE BADZLINA ADANI

11351201771

Tanggal Sidang :

Periode Wisuda :

Jurusan Teknik Informatika

Fakultas Sains dan Teknologi

Universitas Islam Negeri Sultan Syarif Kasim Riau

ABSTRAK

Saat ini telah banyak media sosial yang digunakan masyarakat dari berbagai negara untuk mendapatkan informasi yang dibutuhkan, salah satunya Twitter. Twitter terdapat berbagai bahasa yang digunakan para penggunanya dalam menuliskan *tweet*. Bahasa yang tergolong mirip yaitu bahasa Indonesia dan bahasa Malaysia. *Tweet* berbahasa Indonesia dan berbahasa Malaysia pada Twitter ini, masih ambigu dikarenakan sulit dalam menentukan apakah *tweet* tergolong bahasa Indonesia atau bukan bahasa Indonesia (bahasa Malaysia), disebabkan kemiripan bahasa antara bahasa Indonesia dan bahasa Malaysia karena masih didalam rumpun yang sama. Penelitian ini, menggunakan 1000 *tweet* untuk dijadikan *dataset* yang kemudian akan dilakukan beberapa tahapan seperti label manual, *preprocessing*, seleksi fitur dan pembelajaran SVM. dari hasil tahapan yang telah dilakukan, diperoleh tingkat akurasi sebesar 99% dengan menggunakan normalisasi, *threshold* dan pembagian data 80% sebagai data latih dan 20% sebagai data uji. Tetapi yang tidak menggunakan normalisasi, namun tetap menggunakan *threshold* dengan pembagian data 80% sebagai data latih dan 20% sebagai data uji, diperoleh tingkat akurasi sebesar 87%.

Kata Kunci : Bahasa Indonesia, Bahasa Malaysia, Fitur, Klasifikasi, *Preprocessing*, Support Vecrtor Machine, *Threshold*, *Tweet*, Twitter