

Abstrak

Dalam *text preprocessing*, *term weighting* merupakan salah satu tahapan yang sangat penting. Tahapan ini dilakukan dengan tujuan untuk memberikan suatu nilai/bobot pada *term* yang terdapat pada suatu dokumen. Bobot yang diberikan terhadap sebuah *term* bergantung kepada metode yang digunakan untuk membobotinya. Dalam *text mining*, terdapat beberapa macam metode pembobotan yang diantaranya adalah TF, TF-IDF, WIDF, TF-IG, dan TF-RF.

Pada Tugas Akhir ini, beberapa metode pembobotan kata yaitu TF-IDF, dan TF-RF diperbandingkan *output*-nya terhadap performansi kategorisasi teks. Adapun beberapa parameter yang dijadikan tolok ukur untuk membandingkan performansi dari kategorisasi teks tersebut adalah *precision*, *recall* dan *f-measure*. Untuk menguji *output* dari hasil pembobotan, digunakan *tool* klasifikasi yaitu LIBSVM sebagai *classifier*-nya, dengan *linear SVM* sebagai kernel SVM -nya.

Berdasarkan hasil pengujian, didapat bahwa metode pembobotan TF-RF memiliki performansi yang lebih baik dibanding metode pembobotan TF-IDF. Secara umum, TF-RF mengungguli TF-IDF di sebagian pengujian yang dilakukan. Kelebihan TF-RF yang memperhitungkan frekuensi kemunculan suatu *term* pada suatu kategori dan menormalisasikannya ke keseluruhan dokumen, membuat metode ini lebih baik dibanding TF-IDF.

Kata Kunci : *text preprocessing*, *term*, *term weighting*, *TF-IDF*, *TF-RF*