

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, dan WIDF.

Pada Tugas Akhir ini, beberapa metode pembobotan kata yaitu TF, TF-IDF, dan WIDF 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 Weka, dengan Naïve Bayes dan Naïve Bayes Updateable sebagai *classifier*-nya.

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

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