

ABSTRAK

Kebutuhan analisis *text mining* sangat diperlukan dalam menangani teks yang tidak terstruktur tersebut. Salah satu kegiatan penting dalam text mining adalah klasifikasi atau kategorisasi teks. Analisis text mining ini dilakukan agar mempermudah kita dalam mengambil informasi atau mengelolah informasi yang begitu banyak dari dunia internet atau digital, salah satu nya dengan melakukan klasifikasi dengan data yang sudah tersedia. Kategorisasi teks memiliki berbagai cara untuk melakukan pendekatan antara lain pendekatan *probabilistic*, *support vector machine*, *artificial neural network*, atau *decision tree classification*. Dalam pembelajaran statistik. Support Vector Machine dipilih karena metode ini memiliki kelebihan dalam bidang klasifikasi dengan bantuan kernel. Pada tugas akhir ini support vector machine akan mengelompokkan berita berdasarkan topik menjadi 3 bagian atau *class* yaitu : pemerintahan, ekonomi dan olahraga. Kernel pada Support Vector Mechine akan di kombinasikan dengan *stopword*, *tokenisasi*, *tf-idf*, *chi-square* diharapkan memudahkan untuk mengenali berita tersebut tergolong masuk ke dalam kelas topik yang seharusnya.

Dengan trik kernel dan bantuan metode pembobotan, Dokumen Frekuensi, *Chi square* diharapkan dapat membantu klasifikasi teks dengan baik yang non linear serta mampu meningkatkan akurasi, dengan demikian klasifikasi dengan metode support vektor machine dapat akurasi tertinggi dengan kombinasi *stopword*, *tokenizing*, *term frequency* & *chi-square* 47,43 %.

Kata Kunci : *text mining*, *support vector machine*, *tf-idf*, *chi square*, *stopword*, *tokenisasi*.