

Abstrak

Di dunia internet, informasi selalu bertambah jumlah dan ragamnya sehingga kurang terkelola dan terorganisir yang mengakibatkan sulit dalam pengaksesan dan tidak sedikit informasi penting yang terbuang. Dan salah satu contoh kasusnya adalah informasi dalam bentuk artikel berita berbahasa Indonesia. Melalui proses *text mining* dengan metode kategorisasi, kumpulan artikel yang ada dapat terorganisir dengan baik melalui pengelompokan berdasar topik dari setiap isinya sehingga *user* mudah mendapatkan informasi yang diinginkan.

RIPPER merupakan metode kategorisasi yang didasarkan pada konteks, dimana konteks merupakan kehadiran ataupun ketidakhadiran suatu kata dengan kata lain. Tugas akhir ini mengimplementasikan *RIPPER* pada kumpulan artikel berita berbahasa Indonesia dari beberapa sumber berita yang berasal dari *web* dan bersifat *offline*. Untuk menganalisa performansi dari *RIPPER*, penulis membandingkan dengan *Naive Bayes* yang merupakan metode klasifikasi yang nonkonteks pada data bersih dan *noise*.

Berdasar hasil pengujian secara keseluruhan *RIPPER* memiliki performansi yang lebih rendah dari *Naive Bayes* dan lebih lambat dalam pembangunan model dalam data bersih ataupun *noise*. Hal tersebut menyatakan bahwa dengan jumlah data yang digunakan, metode klasifikasi yang memperhatikan konteks kurang tepat dalam mengklasifikasi dokumen dibanding dengan yang tidak memperhatikan konteks. Dan dikarenakan memperhatikan konteks dari setiap isi dokumen sehingga waktu pembangunan *rule* metode klasifikasi *context-sensitive* lebih lama dibandingkan yang *insensitive*.

Kata kunci : *text mining*, metode kategorisasi, *context-sensitive*, *insensitive*, *RIPPER*, *Naive Bayes*