1. Pendahuluan

Latar Belakang

Belakangan ini internet tumbuh dengan sangat cepat. Informasi di internet menjadi sangat banyak. Khususnya di bidang media informasi di Indonesia sekiranya ada 43 ribu portal berita online, dimana portal berita yang telah terverifikasi oleh menkominfo jumlahnya kurang dari 100 [1]. Dengan relatif mudahnya membuat artikel dan biaya distribusi yang lebih murah menjadikan artikel di media online dapat terus bertambah dengan jumlah yang tak terbatas [2]. Mengingat banyaknya jumlah artikel, kategorisasi artikel berita otomatis penting dilakukan, karena minat pembaca terhadap kategori artikel berbeda-beda. Pembaca dapat menyukai kategori kriminal, politik, bisnis, olahraga dan lain-lain. Jika dalam satu artikel terdapat banyak kata yang berelasi terhadap beberapa kategori yang berbeda, ini akan menyulitkan dalam pemilihan kategori. Dengan adanya sistem ini diharapkan dapat menemukan kategori yang tepat terhadap suatu artikel. Sudah ada beberapa peneliti yang melakukan klasifikasi artikel berita otomatis bahasa Indonesia, diantaranya [4, 6]. Kedua paper tersebut menggunakan metode naïve bayes dalam melakukan klasifikasi. Ada beberapa tahapan dalam melakukan kategorisasi otomatis. Tahapan tersebut terdiri dari preprocessing, feature selection, dan klasifikasi [6]. Preprosessing adalah pengelolahan *raw data* agar siap diklasifikasi. menyiapkan teks berita agar siap di proses Sementara *Feature Selection* adalah langkah yang digunakan untuk memilih sebagian fitur untuk digunakan ketika membangun klasifikasi otomatis untuk kategorisasi text [9].

Berdasarkan paper sebelumnya[7, 8] feature selection berpengaruh terhadap akurasi dalam klasifikasi document. Paper [8] membuktikan bahwa penggunaan feature selection dengan metode Information Gain dapat meningkatkan akurasi 2,9% dibandingkan tanpa menggunakan feature selection. Sedangkan paper [7] melakukan analisis menggunakan feature selection dengan beberapa metode yang dilakukan yaitu *Gini Index* (GI) dan *Distinguishing Feature Selector* (DFS). Dari hasil evaluasi diketahui bahwa DFS menghasilkan akurasi akurasi yang lebih baik dibandingkan dengan GI. Pada TA ini akan di analisa beberapa metode feature selection untuk kategorisasi artikel berita otomatis.

Topik dan Batasannya

Klasifikasi berita dilakukan untuk menempatkan artikel berita pada kategori yang tepat. Sebelum melakukan klasifikasi setiap artikel dibagi kata demi kata yang disebut token, dioptimasi dan diseleksi dalam *preprocessing* sehingga terdapat kumpulan token yang siap untuk diklasifikasi. Tetapi kumpulan token tersebut bisa diseleksi lagi dengan metode *Feature Selection* untuk mencari hasil akurasi yang tinggi dan meningkatkan klasifikasi. Ada berbagai macam metode feature selection yang disini akan diidentifikasi pengaruhnya terhadap akurasi dalam peroses klasifikasi.

Artikel-artikel yang digunakan dalam tugas akhir ini adalah artikel dari berita kompas yang kategorinya dibagi menjadi lima yaitu bola, ekonomi, megapolitan, nasional dan otomotif . Data yang digunakan berjumlah 1125 dengan 125 data *training* dan 1000 data *testing*.

Tujuan

Tujuan dari tugas akhir ini adalah untuk mengidentifikasi teknik *Feature Selection* yang cukup baik untuk klasifikasi artikel berita. Pada eksperimen tugas akhir ini juga dilakukan untuk mengetahui jumlah fitur yang tepat tanpa menggunakan seluruh fitur agar mendapatkan performansi yang lebih tinggi.

Organisasi Tulisan

Laporan tugas akhir ini terdiri dari lima bagian. Pada bagian pertama merupakan pendahuluan. Pada bagian kedua dijelaskan penelitian-penelitikan yang dilakukan sebelumnya yang terkait dengan penelitian. Pada bagian ketiga dijelaskan tentang proses sistem yang dibangun, bagian keempat dijelaskan mengenai evaluasi dari penelitian yang telah dilakukan. Terakhir terdapat kesimpulan dari penelitian.