

1. Pendahuluan

Latar Belakang

Salah satu kegiatan utama dalam proses belajar dan mengajar di sekolah adalah membaca. Pada usia sekolah dasar yaitu antara 6 - 12 tahun, anak sudah mempunyai minat pada aktivitas tertentu, termasuk membaca. Tingginya minat baca para anak di sekolah saat jam pelajaran maupun di luar pelajaran bisa menjadi indikasi tingkat pemahaman dan penguasaan materi pelajaran yang didapatkan.

Buku pelajaran atau buku cerita menjadi sumber pengetahuan yang teramat penting karena dapat berpengaruh langsung kepada pembacanya. Ketika buku itu dibaca, dalam konteks ini adalah anak, tentu akan berusaha memaknai setiap kata dan kalimat yang terdapat di dalamnya. Jika kata dan kalimat itu menguraikan sebuah pengetahuan, tentu sang anak akan berusaha memahaminya. Maka, jadilah ia sebagai pribadi pintar yang berpengetahuan. Namun, apa jadinya jika buku itu berisi kata vulgar dan kalimat tak senonoh.

Dalam diskusi pada sebuah situs forum diskusi daring, banyak buku yang beredar di tahun 2013 dan 2014 memuat konten yang tidak sesuai untuk anak. Sebagai contoh, pada Juli 2013, ditemukan cerita yang memuat cerita tak senonoh dalam buku pelajaran Bahasa Indonesia untuk SD dan MI Kelas VI di halaman 57-60[1]. Pada tahun 2017 masih ditemukan kesalahan konten yang sama pada buku cerita berjudul Si Kabayan Super Kocak oleh penerbit Witarasa[2].

Berdasarkan masalah tersebut, maka pada penelitian tugas akhir ini dibangun sebuah sistem untuk mendeteksi teks cerita yang diambil dari beberapa buku cerita dan buku pelajaran tingkat sekolah dasar (SD) tahun 2011 hingga 2019 (khususnya buku pelajaran yang digunakan saat ini yaitu buku pelajaran tematik Kurikulum 2013 revisi tahun 2017). Teks cerita tersebut dipecah menjadi potongan paragraf yang dijadikan sebagai data set dan diklasifikasikan ke dalam kelas kata yang mengandung makna gereflektif atau non gereflektif. Makna gereflektif merupakan makna yang berhubungan dengan kata atau ungkapan yang tabu. Dalam penelitian ini, bagi anak SD adalah konten dewasa berupa kata-kata vulgar dan atau cerita tak senonoh.

Penelitian yang mengangkat masalah deteksi konten gereflektif pada teks cerita anak belum pernah dilakukan sebelumnya. Namun, terdapat penelitian serupa mengenai klasifikasi *hate speech* berbahasa Indonesia menggunakan Naive Bayes[3], yang menunjukkan bahwa klasifikasi akan menghasilkan performansi yang lebih baik jika menambahkan seleksi fitur *Information Gain* dengan normalisasi.

Terdapat juga beberapa penelitian mengenai deteksi *hoax* pada berita berbahasa Indonesia dengan menerapkan algoritma klasifikasi Naive Bayes[4] dengan menggunakan metrik pengujian presisi, *recall*, dan *f-measure*, dengan hasil berikut 0,91; 1; dan 0,95. Kemudian pada penelitian[5][6], sebanyak 600 data digunakan dan mendapatkan akurasi sebesar 82,6%[5] dan 88%[6], serta menyimpulkan keakuratan yang didapatkan tergantung dari data latih serta tahap *preprocessing*.

Metode Naive Bayes Classifier (NBC) terbukti memiliki akurasi yang tinggi dan juga memiliki waktu komputasi yang cepat jika diimplementasikan untuk menyelesaikan sejumlah permasalahan yang ada pada beberapa penelitian sebelumnya.

Berlandaskan beberapa penelitian sebelumnya, **maka penelitian tugas akhir ini menggunakan metode klasifikasi NBC yang merupakan metode sederhana untuk mengklasifikasikan teks berbasis probabilistic yang memiliki performansi yang tinggi serta dapat membangun model klasifikasi dengan data latih yang kecil.** Sehingga penelitian tugas akhir ini diharapkan dapat membantu mengetahui hasil performansi sistem yang dibangun dalam melakukan klasifikasi konten gereflektif pada cerita anak.

Topik dan batasannya

Rumusan masalah pada penelitian tugas akhir ini adalah teks cerita anak SD sejak tahun 2011 ditemukan memuat konten bermakna gereflektif. Makna gereflektif merupakan makna yang berhubungan dengan kata atau ungkapan yang tabu berupa kata-kata vulgar dan atau cerita tak senonoh. Dari permasalahan tersebut, pertanyaan penelitian pada penelitian ini adalah **“Bagaimana membangun sistem dengan performansi yang baik untuk memprediksi konten gereflektif pada cerita anak?”**

Berdasarkan latar belakang di atas, pada tugas akhir ini, telah dilakukan studi dan implementasi sebuah metode untuk mendeteksi adanya konten gereflektif pada teks cerita anak-anak. Metode klasifikasi yang digunakan adalah Naive Bayes Classifier.

Batasan masalah dalam penelitian tugas akhir ini adalah teks cerita yang dijadikan data set adalah teks berbahasa Indonesia untuk anak SD (berumur 6 – 12 tahun). Pada penelitian ini, teks cerita yang mengandung makna gereflektif relatif sedikit sehingga dilakukan pemecahan teks cerita menjadi potongan paragraf dan juga dianggap bahwa beberapa paragraf pada teks cerita tersebut sudah dapat merefleksikan adanya konten gereflektif. Maka dari itu data set adalah potongan paragraf dari teks cerita yang diambil dari buku cerita dan buku pelajaran

yang paling banyak memuat teks cerita wacana yaitu buku pelajaran tematik Kurikulum 2013 revisi tahun 2017 kelas IV tema 8 dengan judul Daerah Tempat Tinggalku.