



Deteksi Konten Gereflektor pada Cerita Anak Menggunakan Naïve Bayes Classifier

Mayya Tania Wewenggang¹, Dana Sulistiyo Kusumo², Widi Astuti^{1,*}

¹ Fakultas Informatika, Informatika, Universitas Telkom, Bandung, Indonesia

² Fakultas Informatika, Rekayasa Perangkat Lunak, Universitas Telkom, Bandung, Indonesia

Email: ¹wtmayya@student.telkomuniversity.ac.id, ²danakusumo@telkomuniversity.ac.id,

^{3,*}astutiwidi@telkomuniversity.ac.id

Email Penulis Korespondensi: astutiwidi@telkomuniversity.ac.id

Abstrak—Buku pelajaran dan buku cerita adalah salah satu sumber pengetahuan bagi anak-anak. Ketika buku itu dibaca, tentu anak akan berusaha memaknai setiap kata dan kalimat di dalamnya. Namun, muncul masalah ketika ditemukan kesalahan konten pada buku tersebut. Konten yang dimaksud yaitu kata-kata vulgar dan kalimat tak senonoh. Bagi anak-anak di tingkat SD konten tersebut menjadi hal yang bermakna gereflektor (tidak boleh atau tabu). Berdasarkan masalah tersebut, maka dilakukan penelitian terhadap cerita anak yang diambil dari buku fiksi dan buku pelajaran. Penelitian ini dilakukan dengan membangun sistem untuk mendeteksi konten gereflektor pada teks cerita yang dijadikan sebagai data set. Penelitian dilakukan dengan membangun model menggunakan algoritma klasifikasi teks Naïve Bayes Classifier (NBC). Sistem kemudian dievaluasi dengan dua skenario menggunakan metrik pengukuran akurasi, presisi, dan *recall* karena karakteristik data set yang *imbalanced* atau tidak rata dengan jumlah data pada kelas negatif lebih besar dibandingkan dengan data pada kelas positif. Berdasarkan hasil evaluasi, skenario pengujian menghasilkan rata-rata presisi yang tinggi yaitu 99,01%, sedangkan untuk nilai recall memiliki rata-rata di atas 50%. Dari kedua nilai tersebut, dapat disimpulkan bahwa model yang dibangun oleh sistem belum dapat mendeteksi kelas dengan baik, tetapi sangat terpercaya ketika sistem dapat membangun model dengan baik.

Kata Kunci: Cerita Anak, Data Imbalanced, Konten Gereflektor, Klasifikasi Teks, Naïve Bayes Classifier

Abstract—Textbooks and storybooks are the ones used as a source of knowledge. When children read a book, they will try to interpret each word and sentence in it. However, it will be a problem if the book contains vulgar words and indecent sentences. For children at the elementary school level, it is not allowed. For this research, we called that content as gereflektor content. Based on these problems, this research was conducted by building a system to detect gereflektor content in the text of the child's stories that were used as a data set. A system is built by using Naïve Bayes Classifier (NBC) and then evaluated in two scenarios using accuracy, precision, and recall metrics because the characteristics of the data set are imbalanced with the amount of data in the negative class are greater than the data in the positive class. From evaluation results, test scenario produced a high average precision of 99.01%, whereas the recall value has an average of above 50%. From these two values, it can be concluded that the model built by the system has not detected the class properly, but highly trusted when it does.

Keywords: Children Story, Imbalanced Data, Gereflektor Content, Text Classification, Naïve Bayes Classifier

1. PENDAHULUAN

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 Witarisa[2].

Berdasarkan masalah tersebut, maka pada penelitian 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 gereflektor atau non gereflektor. Makna gereflektor (Belanda: *gereflecteerde betekenis*) merupakan makna yang berhubungan dengan kata atau ungkapan yang tabu yang muncul akibat reaksi kita terhadap makna lain[3]. Dalam penelitian ini, bagi anak SD adalah konten dewasa berupa kata-kata vulgar dan atau cerita tak senonoh.

Penelitian yang mengangkat masalah deteksi konten gereflektor pada teks cerita anak belum pernah dilakukan sebelumnya. Namun, terdapat penelitian serupa mengenai klasifikasi hate speech berbahasa Indonesia



menggunakan Naive Bayes[4], yang menunjukkan bahwa klasifikasi akan menghasilkan performansi yang lebih baik jika menambahkan seleksi fitur Information Gain dengan normalisasi.

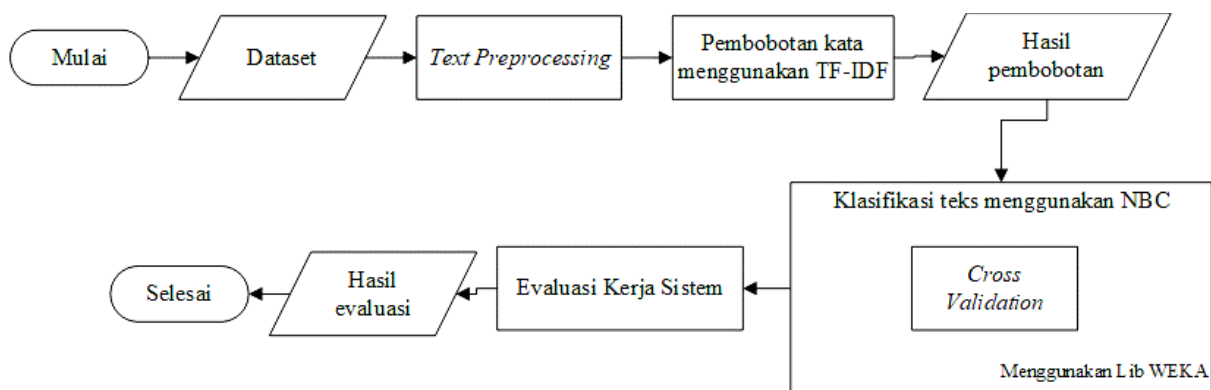
Terdapat juga beberapa penelitian mengenai deteksi hoax pada berita berbahasa Indonesia dengan menerapkan algoritma klasifikasi Naive Bayes[5] dengan menggunakan metrik pengujian presisi, recall, dan f-measure, dengan hasil berikut 0,91; 1; dan 0,95. Kemudian pada penelitian[6], [7], sebanyak 600 data digunakan dan mendapatkan akurasi sebesar 82,6%[6] dan 88%[7], serta menyimpulkan keakuratan yang didapatkan tergantung dari data latih serta tahap preprocessing. Selain itu, terdapat penelitian, yaitu [8], [9], dan [10] yang menyimpulkan bahwa kinerja Naive Bayes Classifier terhadap metode lain dalam beberapa kasus klasifikasi teks.

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 ini menggunakan metode klasifikasi NBC yang merupakan metode sederhana untuk mengklasifikasikan teks berbasis probabilitistik yang memiliki performansi yang tinggi serta dapat membangun model klasifikasi dengan data latih yang kecil.

2. METODE PENELITIAN

Sistem yang dibangun berfungsi untuk mendeteksi konten gereflektor pada teks cerita yang dimuat di dalam buku cerita anak dan buku pelajaran SD dari berbagai sumber sejak tahun 2011 hingga buku pelajaran tematik SD Kurikulum 2013 revisi tahun 2017. Sistem ini mengelompokkan potongan paragraf (yang dijadikan sebagai data set) ke dalam dua kelas, yaitu kelas positif (gereflektor) dan kelas negatif (non-gereflektor) dengan menggunakan algoritma Naive Bayes Classifier sebagai teknik klasifikasinya. Alur kerja sistem yang telah dibangun digambarkan pada Gambar 1.



Gambar 1. Gambaran Umum Sistem

2.1 Dataset

Pada penelitian ini, data set adalah potongan paragraf yang diambil dari teks cerita wacana berbahasa Indonesia, seperti cerita pendek dan dongeng. Teks cerita tersebut diambil dalam beberapa buku pelajaran SD dan buku cerita anak dari beberapa sumber.

Hal pertama yang dilakukan adalah mengumpulkan berita dari beberapa situs resmi berita daring yang menerbitkan berita tentang beredarnya buku pelajaran maupun buku bacaan anak berisi teks cerita yang memuat konten gereflektor sehingga meresahkan para orang tua. Konten gereflektor pada penelitian ini adalah kata-kata vulgar dan atau cerita tak senonoh.

Terkumpul sebanyak enam cerita yang memuat konten gereflektor dari berbagai macam sumber, sehingga teks cerita yang mengandung makna gereflektor relatif sedikit dan mengakibatkan terjadinya imbalance data. Rincian buku yang digunakan adalah sebagai berikut, dua buah buku pelajaran tahun 2011, satu buah buku pelajaran tahun 2013, satu buah buku pelajaran tahun 2015, dua buah buku cerita anak tahun 2017, dan buku pelajaran tematik Kurikulum 2013 revisi tahun 2017 kelas IV tema 8 dengan judul Daerah Tempat Tinggalku.

Setelah dipecah menjadi potongan paragraf, data set untuk kelas gereflektor sebanyak 14 data. Keempat belas data tersebut dianggap sudah dapat merefleksikan adanya konten gereflektor. Sedangkan untuk kelas non gereflektor, sebanyak 107 buah data potongan paragraf telah diambil dari teks cerita wacana pada buku pelajaran tematik Kurikulum 2013 revisi tahun 2017 kelas IV tema 8 dengan judul Daerah Tempat Tinggalku.

2.2 Text Pre-processing

Sistem dimulai dengan melakukan *load* data set. Sistem mengambil data set dari dua *folder* yang sudah dibentuk sebelumnya untuk masing-masing kelas, kemudian ditampung ke obyek *array* dalam sistem.



Selanjutnya, sistem akan melakukan *preprocessing* terhadap data set yang sudah ditampung. Tahap *preprocessing* dimulai dengan *tokenization*. Data set tersebut akan dipotong menjadi kata per kata. Setelah itu, dilakukan tahap *case folding* yang merupakan proses menyeragamkan huruf dengan mengubah semua huruf besar yang ada menjadi huruf kecil. Kemudian karakter selain huruf berupa simbol atau angka akan dihilangkan dan dianggap sebagai delimiter atau pembatas.

Tahap *preprocessing* dilanjutkan dengan menghilangkan *stopword*. Proses ini dilakukan untuk mengeliminasi kata yang tidak penting atau tidak mempunyai makna. Sistem akan melakukan pengecekan kata *stopword*, apabila ada kata yang cocok dengan *list stopwords* yang ada, maka kata tersebut dihapus. Selanjutnya, sistem akan mengolah data dengan proses *stemming*. *Stemming* merupakan tahap merubah kata berimbuhan seperti infiks, prefiks, surfixs, dan sebagainya, menjadi kata dasar. Pada penelitian ini, algoritma *stemming* yang digunakan adalah algoritma Nazief dan Adriani karena seluruh teks dalam Bahasa Indonesia.

2.3 Ekstraksi Fitur Term Frequency-Inverse Document Frequency (TF-IDF)

Setelah tahap *preprocessing* selesai, maka akan dilakukan tahap pembobotan kata. Tahap ini dilakukan untuk menghitung nilai kontribusi (bobot) suatu kata pada data set. Pada tahap ini, algoritma yang digunakan adalah *Term Frequency-Inverse Document Frequency* (selanjutnya akan disebut TF-IDF). Proses perhitungan ini dimulai dengan menghitung masing-masing nilai *Term Frequency* (TF) dan *Inverse Document Frequency* (IDF). Nilai TF mengasumsikan setiap kata (*term*) memiliki proporsi kepentingan sesuai dengan jumlah munculnya *term* tersebut dalam data set yang ada. Dengan perhitungan ini, bobot suatu *term* pada suatu data (dokumen) adalah sama dengan jumlah munculnya *term* tersebut pada dokumen tersebut[11]. Sedangkan IDF adalah nilai seberapa pengaruhnya suatu *term* terhadap suatu dokumen. Untuk mendapatkan nilai TF-IDF, hasil dari nilai TF dan IDF akan dikalikan. Hasil dari perhitungan TF-IDF akan disimpan ke dalam sebuah *file* agar dapat diproses menggunakan *tools classifier* WEKA dengan menyesuaikan aturan data yang dapat diterima oleh WEKA.

2.4 Naïve Bayes Classifier

Proses selanjutnya adalah tahap klasifikasi menggunakan NBC. Proses ini dilakukan dengan memuat data hasil perhitungan TF-IDF, lalu membagi data menggunakan teknik *K-fold Cross-Validation* (*K-CV*). Teknik *K-CV* berfungsi untuk membagi data set menjadi sejumlah *K*-buah partisi secara acak[12], [13]. Sistem akan melakukan sebanyak *K*-kali eksperimen, di mana masing-masing eksperimen menggunakan data partisi ke-*K* sebagai data uji dan memanfaatkan sisa partisi lainnya sebagai data latihan[12].

Proses klasifikasi dilakukan sebanyak *K*-iterasi yang dimulai dari proses pelatihan yaitu sistem menghitung dengan menggunakan persamaan (6) terhadap masing-masing data dan membentuk suatu model. Model yang didapatkan akan digunakan untuk klasifikasi data uji. Pada tahap pelatihan dilakukan proses analisis terhadap sampel dokumen berupa pemilihan *vocabulary*, yaitu kata yang mungkin muncul dalam koleksi dokumen sampel yang sedapat mungkin merepresentasikan dokumen. Selanjutnya adalah penentuan probabilitas *prior* bagi tiap kategori berdasarkan sampel dokumen. Pada tahap klasifikasi ditentukan nilai kategori dari suatu dokumen berdasarkan term yang muncul dalam dokumen yang diklasifikasi.

Lebih konkritnya jika diasumsikan dimiliki dokumen $X(w, \text{frekuensi } w)$ dan untuk setiap label y , dibuat model probabilistik $P = (X|Y = y)$ terhadap dokumen di kelas y . Selanjutnya klasifikasi dokumen adalah mencari nilai maksimum dari:

$$\hat{y} = \underset{y}{\operatorname{argmax}} P(X|y) * P(y) \tag{1}$$

Kemudian dengan mengasumsikan bahwa urutan kata-kata dalam dokumen X tidak berpengaruh dan kata-kata muncul secara independen satu sama lain, sehingga didapatkan persamaan berikut:

$$P(w_1, \dots, w_n|Y = y) = \prod_{i=1}^n P(w_i|Y = y) \tag{2}$$

$$P(W = w_i|Y = y) = \frac{\text{count}(W=w_i \& Y=y)}{\text{count}(Y=y)} \tag{3}$$

Namun kedua persamaan tersebut dapat menimbulkan masalah sehingga untuk menjaga stabilitas numerik, diterapkan persamaan logaritma terhadap persamaan (2), serta menggunakan nilai m dan p yang merupakan *Dirichlet prior* untuk distribusi multinomial terhadap persamaan (3) sehingga didapatkan:

$$\log(P(w_1, \dots, w_n|Y = y)) = \sum_{i=1}^n \log(P(w_i|Y = y)) \tag{4}$$

$$P(W = w_i|Y = y) = \frac{\text{count}(W=w_i \& Y=y) + mp}{\text{count}(Y=y)} \tag{5}$$

Kemudian dengan menyatukan persamaan (4) dan persamaan (5), maka untuk mengklasifikasikan suatu dokumen dipilih nilai tertinggi dari persamaan:

$$\text{score}(y, w_1, \dots, w_n) = \log \frac{\text{count}[y]}{\text{count}} + \sum_{i=1}^n \log \frac{\text{count}[w_i|y] + mp}{\text{count}[y] + m} \tag{6}$$



2.5 Evaluasi Sistem

Evaluasi dilakukan dengan tujuan untuk mengetahui akurasi hasil klasifikasi data menggunakan metode Naïve Bayes Classifier terhadap data uji. Dengan evaluasi dapat diketahui tingkat akurasi dari penggunaan seleksi fitur dan klasifikasi pada data uji. Evaluasi dalam penelitian ini dilakukan pada data uji terhadap fungsi klasifikasi yang telah dilatih. Hasil pengujian tersebut berupa kelas prediksi yang akan dibandingkan dengan kelas aktual pada data uji dengan *confusion matrix*.

Tabel 1. Confusion Matrix

Class	Actual Positive	Actual Negative
Predicted Positive	TP	FP
Predicted Negative	FN	TN

Dengan TP (*True Positive*) adalah kelas yang diprediksi oleh sistem positif dan sesuai dengan data aslinya, yaitu positif. TN (*True Negative*) adalah kelas yang diprediksi oleh sistem sebagai data negatif dan sesuai dengan data aslinya yaitu negatif. FP (*False Positive*) adalah kelas yang diprediksi oleh sistem positif, tetapi data aslinya memiliki nilai negatif. FN (*False Negative*) adalah kelas yang diprediksi oleh sistem negatif, namun data aslinya positif.

Pengukuran performansi sistem dapat dilakukan dengan menghitung nilai akurasi, presisi, dan *recall*[14]. Nilai akurasi dapat menunjukkan tingkat keberhasilan sistem dalam melakukan klasifikasi. Akurasi dapat dihitung dengan persamaan berikut:

$$Akurasi = \frac{TP+TN}{TP+FP+FN+TN} \tag{7}$$

Presisi (P) adalah ukuran banyaknya dokumen yang ditemukan relevan, dinyatakan dalam pecahan sebagai berikut;

$$Precision = \frac{TP}{TP+FP} \tag{8}$$

Sedangkan *recall* (R) adalah ukuran banyaknya dokumen yang relevan dapat ditemukan kembali, dinyatakan dalam pecahan sebagai berikut;

$$Recall = \frac{TP}{TP+FN} \tag{9}$$

Namun, berdasarkan kondisi *imbalance data* yang terjadi yaitu jumlah data kelas negatif (non-gereflektor) jauh lebih banyak dibandingkan dengan data kelas positif (gereflektor) maka evaluasi performansi sistem yang akan dianalisis untuk mengukur performansi sistem adalah metrik presisi dan *recall*.

3. HASIL DAN PEMBAHASAN

Pada penelitian ini, dilakukan dua buah skenario pengujian, yaitu mengubah nilai *K* dan melakukan perubahan jumlah data kelas negatif. Pada kedua skenario ini, akan diterapkan teknik *random undersampling* untuk menangani sifat data set yang *imbalanced*[15].

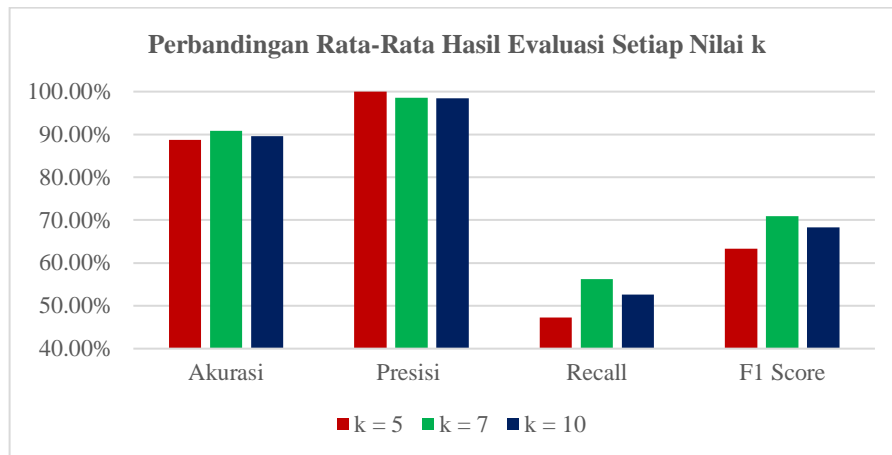
3.1 Hasil Pengujian dan Analisis Skenario I: Perubahan Nilai K

Skenario pertama bertujuan untuk mengetahui penggunaan nilai *K* yang paling optimal menggunakan teknik *K-CV*. Pada skenario ini, digunakan tiga nilai *K* yang berbeda, yaitu *K* = 5, *K* = 7, dan *K* = 10. Untuk masing-masing nilai *K*, diterapkan perubahan jumlah data set kelas negatif dengan kelipatan 14. Hal ini dilakukan untuk mengatasi masalah jumlah data set tidak rata (*imbalance data*), yaitu kelas positif lebih kecil dibandingkan dengan kelas negatif dan saat rasio data kelas positif dan negatif seimbang.

Pengujian dimulai dari 14 data kelas negatif, jumlah data tersebut agar jumlah data set kedua kelas sama rata yaitu masing-masing 14 data. Pengujian berikutnya digunakan 28 data, 42 data, 56 data, 70 data, 84 data, 98 data, dan 107 data kelas negatif, sehingga terdapat delapan hasil evaluasi. Setelah itu, kedelapan hasil evaluasi tersebut dihitung rata-ratanya untuk dianalisa. Rata-rata hasil pengujian skenario I dapat dilihat pada Tabel 1 dan Gambar 2.

Tabel 2. Rata-rata Hasil Evaluasi Setiap Nilai K

Nilai K	Akurasi (%)	Presisi (%)	Recall (%)	F1 Score (%)
5	88,67	100	47,32	63,31
7	90,89	98,61	56,25	70,98
10	89,61	98,44	52,68	68,28
Rata-rata	89,72	99,01	52,08	67,52

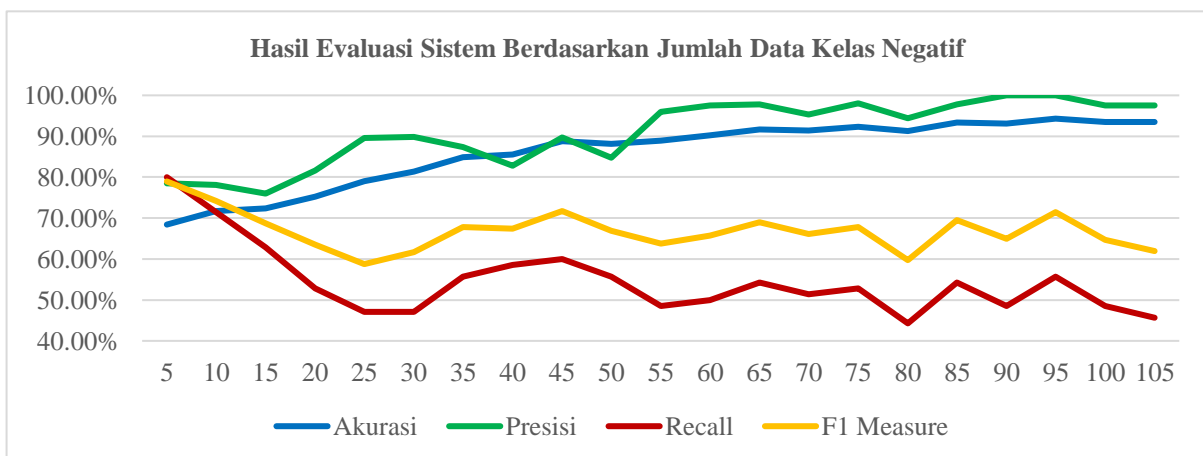


Gambar 2. Grafik Perbandingan Rata-rata Hasil Evaluasi Setiap Nilai K

Pada Gambar 2, nilai presisi yang tertinggi saat nilai $K = 5$, yaitu sebesar 100%. Hal ini disebabkan karena nilai $FP = 0$ atau tidak ada data negatif yang diprediksi sebagai data positif. Sedangkan ketika nilai $K = 7$ dan $K = 10$, terdapat sebuah data kelas negatif yang salah prediksi. Untuk nilai *recall*, yang terbaik adalah sebesar 56,25% ketika $K = 7$. Untuk pengukuran *F1 Score*, yang terbaik adalah dengan menggunakan $K = 7$ yaitu 70,98%. Dengan ini dapat disimpulkan bahwa ketika nilai $K = 7$, sistem dapat menghasilkan hasil yang paling optimal untuk menangani masalah *imbalance data*, dengan nilai akurasi sebesar 90,89%, nilai presisi sebesar 98,61%, nilai *recall* sebesar 56,25%, dan *F1 Score* sebesar 70,98%.

3.2 Hasil Pengujian dan Analisis Skenario II: Perubahan Jumlah Data Kelas Negatif (Non-gereflektor)

Pada skenario kedua ini, pengujian dilakukan dengan merubah jumlah data kelas negatif dengan menggunakan nilai K yang besarnya sama dengan jumlah data. Tujuannya untuk melihat pengaruh perbandingan data antar kelas terhadap kinerja sistem. Pengujian dilakukan dengan merubah jumlah data negatif dengan kelipatan 5. Dimulai dengan menggunakan lima buah data negatif, sehingga rasio data set positif:negatif adalah 14:5. Kemudian dilakukan penambahan lima data kelas negatif menjadi sepuluh data negatif hingga menggunakan 105 data. Untuk masing-masing perubahan jumlah data negatif, dilakukan sebanyak lima kali percobaan dengan mengacak data negatif yang digunakan. Hasil pengujian skenario II dapat dilihat pada Gambar 3.



Gambar 3. Hasil evaluasi sistem

Pada Gambar 3, dapat dilihat bahwa seiring bertambahnya jumlah data negatif akurasi dan presisi, meningkat sementara *recall* menurun. Berdasarkan grafik presisi yang tinggi dan *recall* yang rendah dapat disimpulkan bahwa model yang dibangun oleh sistem belum dapat mendeteksi kelas dengan baik, tetapi sangat terpercaya ketika sistem dapat membangun model dengan baik, serta perubahan jumlah data kelas negatif mempengaruhi nilai *recall*. Semakin besar data kelas negatif yang membuat data set menjadi *imbalanced*, maka nilai *recall*-nya semakin kecil. Hal ini disebabkan bertambahnya jumlah data positif yang diprediksi sebagai data negatif akibat persentase data latih positif yang semakin kecil.



3.3 Analisis Data Kelas Positif yang Diprediksi sebagai Kelas Negatif

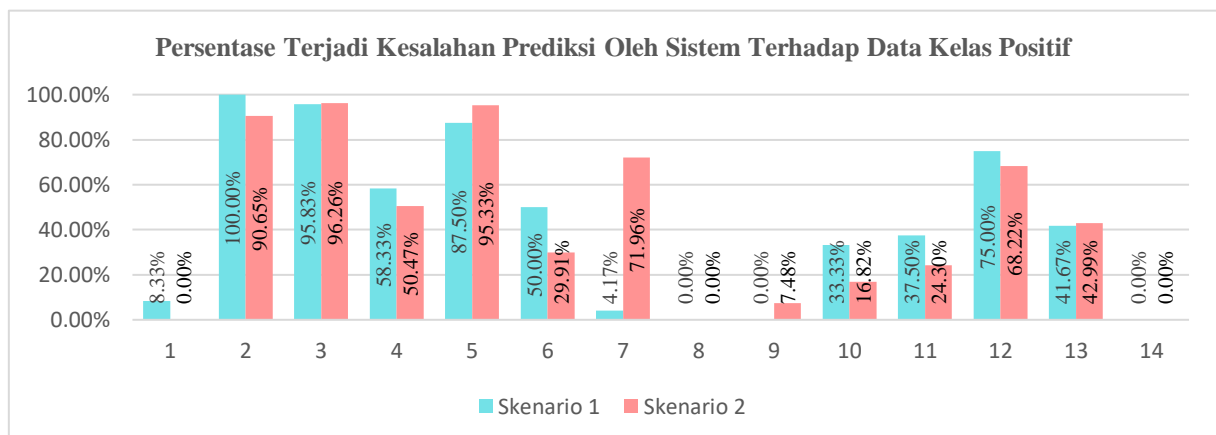
Berdasarkan nilai *recall* yang kecil karena disebabkan oleh nilai FN yang tinggi, maka bagian ini berisi hasil analisis terhadap setiap data kelas positif (gereflektor) yang salah diprediksi oleh sistem. Hal ini bertujuan untuk mengetahui karakter seperti apa yang dimiliki oleh data set kelas positif sehingga diprediksi sebagai data kelas negatif oleh sistem.

Untuk setiap data kelas positif di masing-masing skenario, dilakukan perhitungan persentase *error rate* yaitu saat sistem memprediksi kelas positif sebagai data negatif. Untuk skenario I yang menggunakan tiga percobaan dan tiga nilai *K*, hasil perhitungan persentase dari ketiga percobaan kemudian dirata-ratakan. Hasil perhitungan persentase untuk kedua skenario yang lebih rinci dapat dilihat pada Lampiran 6. Persentase rata-rata data kelas positif yang salah diprediksi oleh sistem pada kedua skenario dapat dilihat pada Tabel 2 dan Gambar 4.

Tabel 2. Persentase Kejadian Kelas Positif yang Diprediksi sebagai Kelas Negatif pada Skenario 1 dan 2

Data set ke-	Skenario 1 (%)	Skenario 2 (%)
1	8,33	0
2	100	90,65
3	95,83	96,26
4	58,33	50,47
5	87,50	95,33
6	50	29,91
7	79,17	71,96
8	0	0
9	4,17	7,48
10	33,33	16,82
11	37,50	24,30
12	75	68,22
13	41,67	42,99
14	0	0

Nilai persentase 0% dan 100% adalah nilai mutlak, yaitu 0% ketika sistem selalu berhasil memprediksi suatu data sebagai kelas positif dan 100% ketika sebuah data kelas positif selalu diprediksi sebagai kelas negatif. Sedangkan rentang nilai persentase di antara 0,01 - 99,99% memiliki jumlah yang berbeda pada kedua skenario. Contohnya rentang nilai 75,01 - 99,99% pada skenario I adalah terjadi sebanyak 7 dari 8 kali percobaan saat sistem tidak tepat memprediksi kelas aktual data kelas positif. Sedangkan pada skenario II, untuk rentang nilai yang sama berarti sebanyak 82 sampai 106 dari 107 kali percobaan.



Gambar 4. Grafik persentase kesalahan prediksi

Dalam konteks ini ada dua dugaan penyebab terjadinya kesalahan saat memprediksi suatu data kelas positif. Pertama adalah panjang (jumlah kata) data tersebut dan yang kedua adalah kata yang terkandung dalam suatu data. Pada Gambar 4 dapat dilihat bahwa ada enam buah data yang persentase salah prediksinya di atas 50%, yaitu data ke-2, 3, 4, 5, 7, dan 12 dengan panjang masing-masing data adalah 75; 40; 44; 85; 88; dan 44 (dapat dilihat pada Lampiran 2). Jika diambil kesimpulan bahwa data set yang memiliki jumlah kata dengan rentang 40 – 88 kata sangat mungkin sistem akan memprediksi data kelas positif sebagai kelas negatif. Hal ini tidak dapat dibuktikan karena data yang lain memiliki jumlah kata yang serupa namun berhasil diprediksi oleh sistem, sehingga dugaan pertama salah. Sedangkan untuk dugaan yang kedua, pada data ke-2 dan ke-3 terdapat kata pembunuh dan membunuh. Ketika melalui proses *stemming*, hasilnya adalah mbunuh. Sangat mungkin kata tersebut adalah penyebabnya, namun hal ini juga tidak dapat disimpulkan sebagai penyebab kesalahan prediksi karena pada data



set ke-1 terdapat juga kata pembunuh yang setelah *distemming* menghasilkan kata mbunuh. Berdasarkan hasil analisis kedua dugaan tersebut, tidak ada karakteristik khusus untuk suatu data kelas positif yang mungkin akan sering atau selalu diprediksi sebagai kelas negatif.

4. KESIMPULAN

Berdasarkan hasil evaluasi yang telah dilakukan, penggunaan data set yang tidak rata (*imbalance data*) yaitu kelas positif lebih kecil dibandingkan dengan kelas negatif maka dianalisis hasil evaluasi nilai presisi dan *recall*.

Pada skenario pertama pengujian menghasilkan rata-rata nilai presisi yang tinggi, yaitu 99,01% karena nilai FP yang sangat kecil. Sedangkan untuk nilai *recall* memiliki hasil rata-rata di atas 50%. Hal ini disebabkan oleh nilai FN yang besar, yang artinya 7 dari 14 data kelas positif masih salah diprediksi oleh sistem.

Pada skenario kedua nilai presisi yang tinggi dan nilai *recall* yang rendah menyimpulkan bahwa model yang dibangun oleh sistem tidak dapat mendeteksi kelas dengan baik, tetapi sangat terpercaya ketika sistem dapat membangun model dengan baik.

Kejadian data kelas positif yang diprediksi sebagai kelas negatif dapat dipengaruhi oleh beberapa faktor. Namun, dalam penelitian yang telah dilakukan masih belum dapat disimpulkan penyebab utama kesalahan tersebut terjadi. Dalam penelitian ini, penulis belum bisa menyimpulkan karakteristik khusus data yang masih salah diprediksi oleh sistem.

Semakin banyak kelas positif yang benar diprediksi oleh sistem, maka nilai FN semakin kecil dan nilai *recall* akan semakin tinggi. Sama halnya ketika kelas negatif benar diprediksi oleh sistem, maka nilai FP semakin kecil sehingga nilai presisi akan semakin tinggi. Ketika kedua nilai presisi dan *recall* tinggi, maka sistem yang digunakan dapat dikatakan berjalan dengan optimal.

Klasifikasi konten gereflektor pada cerita anak akan memberikan hasil yang lebih baik jika penelitian yang akan dilakukan selanjutnya dapat menambahkan data set yang mengandung makna gereflektor sehingga rasio kelas positif dan kelas negatif menjadi sedikit lebih seimbang, serta dengan mengoptimalkan proses *stemming* dengan memperhatikan kamus kata dasar algoritma yang digunakan.

REFERENCES

- [1] Davinof, "6 Buku Pelajaran yang Pernah Bikin Geger Dunia Pendidikan (+PICT)," *Kaskus*, 2014. [Online]. Available: <https://www.kaskus.co.id/thread/5441ba4adc06bd784d8b457a/6-buku-pelajaran-yang-pernah-bikin-geger-dunia-pendidikan-pict/>. [Accessed: 06-Feb-2020].
- [2] N. Purnamasari, "Orang Tua Keluhkan Konten Dewasa di Buku 'Si Kabayan Super Kocak,'" *detiknews*, 2017. [Online]. Available: <https://news.detik.com/berita/d-3547378/orang-tua-keluhkan-konten-dewasa-di-buku-si-kabayan-super-kocak>. [Accessed: 06-Feb-2020].
- [3] M. Pateda, *Semantik leksikal*. Gorontalo: Rineka Cipta, 1996.
- [4] Ivan, Y. A. Sari, and P. P. Adikara, "Klasifikasi Hate Speech Berbahasa Indonesia di Twitter Menggunakan Naive Bayes dan Seleksi Fitur Information Gain dengan Normalisasi Kata," *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 3, no. 5, pp. 4914–4922, 2019.
- [5] K. N. Sani, "Sistem Deteksi Hoax Berita Bahasa Indonesia dengan Menggunakan Algoritma Naive Bayes," Universitas Airlangga, 2018.
- [6] F. Rahutomo, I. Y. R. Pratiwi, and D. M. Ramadhani, "Eksperimen Naive Bayes Pada Deteksi Berita Hoax Berbahasa Indonesia," *J. Penelit. Komun. dan Opini Publik*, vol. 23, no. 1, pp. 1–15, 2019.
- [7] P. D. Utami and R. Sari, "Filtering Hoax Menggunakan Naive Bayes Classifier," *J. MULTINETICS*, vol. 4, no. 1, pp. 57–61, 2018.
- [8] M. A. Rahman and Y. A. Akter, "Topic Classification from Text Using Decision Tree, K-NN and Multinomial Naive Bayes," in *2019 1st International Conference on Advances in Science, Engineering and Robotics Technology (ICASERT)*, 2019, pp. 1–4.
- [9] S. R. Basha, J. K. Rani, J. P. Yadav, and G. R. Kumar, "Impact of feature selection techniques in Text Classification: an experimental study," in *2nd International Conference on Advances in Engineering, Management and Sciences*, 2019, pp. 39–51.
- [10] R. Chatterjee, V. Acharya, K. Prakasha, and R. V. Arjunan, "Text based Machine Learning Using Discriminative Classifiers," *J. Adv. Res. Dyn. Control Syst.*, vol. 11, no. 7, pp. 32–41, 2019.
- [11] T. Tokunaga and I. Makoto, "Text categorization based on weighted inverse document frequency," in *Information Process Society of Japan (SIG-IPSJ)*, 1994, pp. 33–39.
- [12] S. kotagiri Raju and M. R. Murty, "Support Vector Machine with K-fold Cross Validation Model for Software Fault Prediction," *Int. J. Pure Appl. Math.*, vol. 118, no. 20, pp. 321–334, 2018.
- [13] P. Refaailzadeh, L. Tang, and H. Liu, "Cross-Validation," in *Encyclopedia of Database Systems 5*, Boston, MA: Springer, 2009.
- [14] J. Ling, I. P. E. N. Kencana, and T. B. Oka, "Analisis Sentimen Menggunakan Metode Naive Bayes Classifier Dengan Seleksi Fitur Chi Square," *E-Jurnal Mat.*, vol. 3, no. 3, pp. 92–99, 2014.
- [15] S. Mishra, "Handling imbalanced data: SMOTE vs. random undersampling," *Int. Res. J. Eng. Technol.*, vol. 4, no. 8, pp. 317–320, 2017.