

Identifikasi *Cyberbullying* pada Komentar Instagram menggunakan Metode *Lexicon-Based* dan *Naïve Bayes Classifier* (Studi kasus: Pemilihan Presiden Indonesia Tahun 2019)

Rizky Dhian Syarif¹, Anisa Herdiani, S.T., M.T.², Widi Astuti, S.T., M.Kom.³

^{1,2,3}Fakultas Informatika, Universitas Telkom, Bandung

¹kikysyarif@students.telkomuniversity.ac.id, ²anisaherdiani@telkomuniversity.ac.id,

³widiwdu@telkomuniversity.ac.id

Abstrak

Tahun 2019 Indonesia diwarnai dengan semarak demokrasi. Masyarakat menyambut dengan gembira dan antusiasme yang tinggi pada Pemilihan Umum Presiden yang dilaksanakan April 2019. Pilpres ini ramai diperbincangkan di dunia nyata maupun dunia maya, khususnya di media sosial Instagram. Semua orang bebas berpendapat atau beropini tentang masing-masing calon Presiden. Tetapi, yang menjadi persoalan adalah ketika berpendapat tidak berlandaskan etika, sehingga membuat pertentangan antara masing-masing pendukung pasangan calon presiden. Perang komentar yang membully, menjelekkan, atau menjatuhkan lawan mewarnai situasi tersebut. Untuk itu, perlu dilakukan identifikasi *cyberbullying* pada komentar Instagram untuk mengklasifikasikan komentar yang mengandung *cyberbullying* atau *non cyberbullying*. Metode yang digunakan dalam penelitian ini adalah metode berbasis *lexicon* dan metode berbasis *learning* yaitu *naïve bayes classifier*. Proses sistem dimulai dari *text preprocessing* dengan tahapan *cleaning*, *casefolding*, dan *stemming*. Kemudian dilakukan proses klasifikasi menggunakan metode *Lexicon based* dan *naïve bayes classifier*, dan hasil keluaran sistem berupa identifikasi apakah komentar termasuk *cyberbullying* atau *non cyberbullying*. Pada penelitian ini didapatkan hasil performansi dari metode *Lexicon-Based* menghasilkan akurasi sebesar 58%, presisi 52%, recall 75% dan F-score 61%. Sedangkan *naïve bayes classifier* didapatkan akurasi 97%, presisi 94%, recall 100%, dan F1-score 97%.

Kata kunci : *cyberbullying*, *instagram*, *Lexicon-Based*, *naïve bayes classifier*.

Abstract

In 2019 Indonesia was colored with the vibrant democracy. The community welcomed with great enthusiasm and enthusiasm at the Presidential Election held in April 2019. The presidential election was heavily discussed in the real world and cyberspace, specifically on Instagram social media. All people are free to approve or opinion about each candidate for President. However, what is being debated is a compilation that is not based on ethics, thus creating a conflict between each of the supporters of the presidential candidate pair. The war of comments that bully, vilify, or bring down opponents depicts beforehand. For this reason, it is necessary to collect cyberbullying on Instagram comments to classify comments that contain cyberbullying or non-cyberbullying. The method used in this research is the lexicon based method and the Bayes classifier naïve learning method. The system process starts from preprocessing text with cleaning, casefolding, and stemming. Then the classification process is carried out using the Lexicon-based method and the naïve Bayes classifier, and the output of the system involves commenting whether it is cyberbullying or non-cyberbullying. In this study the performance results obtained from the Lexicon-Based method produce an accuracy of 58%, 52% precision, 75% recall and F-score 61%. While Naïve Bayes Classifier obtained 97% accuracy, 94% precision, 100% recall, and F1-score 97%.

Keywords: *cyberbullying*, *instagram*, based on *lexicon*, *naive bayes classifier*.

1. Pendahuluan

Menurut laporan Asosiasi Penyelenggara Jasa Internet Indonesia (APJII), populasi penduduk Indonesia saat ini mencapai 262 juta orang. Lebih dari 50 persen atau sekitar 143 juta orang telah terhubung dengan jaringan internet sepanjang tahun 2017[1]. Survei Ditch The Label lembaga riset asal Inggris yang fokus pada *bullying* menetapkan Instagram menjadi media sosial yang paling banyak terjadi *cyberbullying*[2]. Di Indonesia sendiri sudah diatur dalam undang-undang (UU) tentang Informasi Elektronik (ITE) Nomor 11 Tahun 2008 pasal 27 ayat 3 yang berisi tentang penyebaran nama baik atau penghinaan[3].

Media sosial tentu saja memiliki manfaat yang baik, mulai dari berinteraksi di dunia maya, serta menghibur. Pengguna media sosial yang semakin tinggi juga diikuti dengan lahirnya kebebasan berpendapat. Pengguna media sosial seakan menemukan ruang untuk mengemukakan apa saja yang mereka pikirkan dan rasakan. Namun, tidak dipungkiri juga masalah baru muncul, salah satunya *cyberbullying*. *Cyberbullying* atau perundungan yang banyak terjadi di media sosial menjadikan suatu kekhawatiran tersendiri. *Cyberbullying* dikenal sebagai bentuk “ancaman” atau “serangan” yang dilakukan seseorang terhadap orang lain yang disampaikan melalui pesan elektronik lewat media. *Bullying* terjadi dalam berbagai bentuk diantaranya *bullying* berupa penghinaan, menyebarkan fitnah, mengucilkan orang lain, memperpermalukan, mengintimidasi, serta mengungkapkan rasa kebencian kepada orang lain menggunakan kata kasar [4]. *Cyberbullying* dianggap lebih mudah dilakukan daripada kekerasan konvensional, karena pelaku tidak perlu bertatap muka pada target.

Pemilihan Presiden Indonesia Tahun 2019 merupakan peristiwa yang paling banyak mendapatkan perhatian masyarakat Indonesia. Kedua pasangan calon Presiden, Jokowi-Amin dan Prabowo-Sandi juga melakukan kampanye melalui media sosial. Kampanye di media sosial dinilai efektif dan hemat biaya karena, ide dan gagasan dari masing-masing calon bisa langsung mendapatkan respon dari masyarakat pengguna media sosial, salah satu contoh situs jejaring sosial yang ramai diperbincangkan adalah Instagram. Pemilu ini juga sarat dengan saling *bully* oleh masing-masing pendukung terhadap lawan mereka. Pilpres yang diharapkan berjalan baik, berubah menjadi saling serang, menyudutkan dan menjelekkan satu sama lain. Contoh kasus *cyberbullying* yang terjadi saat debat pemilihan presiden Indonesia yang di *bully* pada sosial media Instagram karena pertanyaan yang dilontarkan Jokowi sebagai capres yang mengusung nomor urut 1 mengenai “mantan narapidana korupsi yang dicalonkan Partai Gerindra sebagai calon legislatif”. Dan pertanyaan tersebut dijawab oleh calon presiden nomor urut 2 yang mengatakan “Mungkin korupsinya nggak seberapa”. Salah satu contoh komentarnya yaitu “bangsat, enak bgt anda ngomong”. Kejadian ini sangat penting dan menarik untuk dianalisis sejauh mana kondisi *cyberbullying* pada media sosial Instagram terhadap akun resmi tim kedua pasangan calon Presiden Indonesia 2019. Pada penelitian ini menerapkan metode *Lexicon Based* dan *naïve bayes classifier* yang tidak dikombinasikan, dan membandingkan hasil akurasi performa pengujian sistem yang dihasilkan dari masing-masing metode sehingga, dapat diketahui performa mana yang lebih unggul dari kedua metode tersebut.

Dalam penelitian ini, algoritma yang digunakan untuk klasifikasi adalah Naïve Bayes Classifier. Jika dilihat dari kompleksitas, *Naïve Bayes Classifier* lebih sederhana dan konvensional daripada algoritma lainnya. Hal ini telah dibuktikan oleh peneliti-peneliti sebelumnya, pada penelitian [5] tentang analisis sentiment terhadap pemerintahan Joko Widodo pada media sosial twitter menggunakan algoritma naïve bayes, hasil penelitian tersebut membuktikan bahwa nilai akurasi sebesar 97%. Untuk nilai akurasi tiap sentimennya yaitu 96% untuk sentimen negatif dan 98% untuk sentimen positif. Oleh karena itu algoritma yang digunakan yaitu algoritma *Naïve Bayes Classifier* untuk mengidentifikasi *cyberbullying* pada komentar Instagram. Serta metode *lexicon based* digunakan sebagai metode pembanding karena *Lexicon based* didasarkan oleh orientasi kontekstual pada jumlah orientasi sentimen untuk setiap kata atau kalimat.

Topik dan Batasannya

Topik yang dibahas dari penelitian ini adalah seputar identifikasi *cyberbullying* pada komentar Instagram, menggunakan akun resmi pasangan calon presiden Indonesia. Batasan masalah pada penelitian ini yaitu data yang digunakan untuk mengidentifikasi adalah data komentar akun resmi tim @jokowi.amin dan @indonesiaadilmakmur yang berbahasa Indonesia. Pada penelitian ini *cyberbullying* yang dimaksud adalah komentar berupa penghinaan, mengucilkan, memperpermalukan, mengintimidasi, dan mengungkapkan rasa kebencian kepada orang lain.

Tujuan

Tujuan utama penelitian ini adalah mengidentifikasi komentar yang mengandung *cyberbullying* pada akun resmi Instagram tim @Jokowi.amin dan @Indonesiaadilmakmur menggunakan metode *Lexicon-Based* dan *naïve bayes classifier*, karena pada Pemilihan Presiden Indonesia Tahun 2019 ini sangat besar peluang terjadinya *cyberbullying*. Oleh karena itu, dengan adanya sistem ini dapat membantu dalam penanganan komentar *cyberbullying* sehingga proses pemilihan presiden Indonesia tahun 2019 berjalan dengan kondusif.

2. Tinjauan Pustaka

2.1 Studi Terkait

Pada penelitian yang dilakukan oleh Restio Afrinza tentang Deteksi *Cyberbullying* pada twitter menggunakan naïve bayes classifier Hasil penelitian tersebut menghasilkan sistem yang dapat mengklasifikasikan sentimen dengan hasil pengujian untuk 3250 komentar *data testing* yaitu akurasi 82,12%, presisi 85,28%, recall 95,22% dan error rate 15,70%[4]. Penelitian yang lain telah dilakukan oleh Sentiaji dan Bachtiar pada [6], bahwa algoritma Naïve Bayes Classifier dapat mengklasifikasikan pendapat dalam bentuk tweets menjadi dua kelas yaitu positif dan negatif akurat.

2.2 Lexicon Based

Lexicon Based merupakan suatu pendekatan yang meliputi frase, bentuk ekspresi, atau konten yang berupa teks yang umumnya terdapat pada obrolan, dialog, post, review, dan lainnya[7]. *Lexicon* merupakan pendekatan yang menggunakan suatu kamus sentimen berisi kata positif dan kata negatif yang dibandingkan dan dicocokkan dengan kata pada kalimat untuk diketahui tingkat polaritasnya[8]. Pada penelitian [7] kamus sentimen juga digunakan untuk Analisis Sentimen *Cyberbullying* pada Komentar Instagram dengan Metode Klasifikasi *Support Vector Machine*.

Berikut adalah kamus yang digunakan dalam penelitian ini dan contoh isi kamusnya:

1. Positif keywords: baik, banyak, bangkit
2. Negatif keywords: bangkrut, banjir, bantah
3. Negation keywords: belum, bukan, tidak
4. Emoticon
5. Kamus gaul: bgmn = bagaimana, bgs = bagus, beud = banget

2.3 Pembobotan

Pembobotan adalah salah satu proses menentukan skor polaritas sentimen pada kalimat opini yang mengandung sentimen. Pembobotan *Lexicon Based* lainnya dilakukan dengan mempertimbangkan skor sentimen dari setiap komentar[7]. Terdapat tiga tahapan dalam proses pembobotan yaitu penentuan polaritas kata, penanganan negasi, dan pemberian skor terhadap kalimat[9]. Pada proses ini memerlukan sentimen *lexicon* atau kamus sentimen sebagai acuan dalam pemberian nilai polaritas suatu kata sentimen. Pencarian nilai sentimen dalam satu kalimat menggunakan rumus 2.1 dan 2.2.

$$S_{positive} = \sum_{i \in T}^{n} positive\ score_i \tag{2.1}$$

$$S_{negative} = \sum_{i \in T}^{n} negative\ score_i \tag{2.2}$$

Dimana (*Spositive*) adalah bobot dari kalimat yang didapatkan melalui penjumlahan n skor polaritas kata positif dan (*Snegative*) adalah bobot dari kalimat yang didapatkan melalui penjumlahan n skor polaritas kata negatif. Bobot pada tiap kata ini yang akan digunakan sebagai acuan untuk melakukan proses perbandingan. Sehingga dalam satu kalimat akan diketahui total jumlah nilai positif (*Spositive*) dan juga nilai negatif (*Snegative*) dari tiap-tiap kata penyusunnya. Dari persamaan nilai sentimen dalam satu kalimat maka diperoleh persamaan 2.3 untuk menentukan orientasi sentimen dengan perbandingan jumlah nilai positif, negative.

$$Sentence\ Sentiment\ \begin{cases} positive\ if\ S_{positive} > S_{negative} \\ negative\ if\ S_{positive} < S_{negative} \end{cases} \tag{2.3}$$

Jika total jumlah nilai positif lebih besar dari jumlah nilai negatif maka kalimat akan berorientasi positif. Namun jika total jumlah nilai positif kurang dari jumlah nilai negatif maka kalimat akan berorientasi negatif.

2.4 Naïve Bayes Classifier

Naive Bayes classifier merupakan suatu algoritma klasifikasi berdasarkan pada probabilitas dengan menjumlahkan frekuensi dan kombinasi nilai dari dataset dengan menggunakan teorema Bayes [4].

$$P(x|y) = \frac{P(y|x) \cdot P(x)}{P(y)} \quad (2.4)$$

H merupakan hipotesis data yang merupakan class spesifik, y adalah data dengan class yang belum diketahui, $P(x|y)$ adalah probabilitas hipotesis x terhadap kondisi y, $P(y|x)$ adalah probabilitas kondisi y terhadap hipotesis x, $P(x)$ adalah probabilitas hipotesis x, $P(y)$ adalah probabilitas y [6]. Pengklasifikasian teks dengan menggunakan nilai probabilitas suatu dokumen d dari semua kelas. \hat{c} adalah perkiraan kelas yang benar. Dapat dihitung dengan persamaan:

$$\hat{c} = \operatorname{argmax}_{c \in C} P(c|d) \quad (2.5)$$

Persamaan 2.4 dapat diganti dengan persamaan 2.5 sehingga mendapatkan persamaan:

$$(2.6) \quad \hat{c} = \operatorname{argmax}_{c \in C} P(c|d) = \operatorname{argmax}_{c \in C} \frac{P(d|c)P(c)}{P(d)}$$

Karena $P(d)$ tidak mengalami perubahan untuk setiap kelas, sehingga probabilitas $P(d)$ diabaikan sehingga diperoleh persamaan [4]:

$$(2.7) \quad \hat{c} = \operatorname{argmax}_{c \in C} P(c|d) = \operatorname{argmax}_{c \in C} P(d|c)P(c)$$

Persamaan 2.7 dapat dituliskan menjadi:

$$\hat{c} = \operatorname{argmax}_{c \in C} P(d|c)P(c) \quad (2.8)$$

Karena pada dokumen d terdapat kata-kata, sehingga dokumen d dapat diganti dengan f_1, f_2, \dots, f_n . Dinyatakan dengan persamaan [4]:

$$\hat{c} = \operatorname{argmax}_{c \in C} P(f_1, f_2, \dots, f_n|c)P(c) \quad (2.9)$$

Naive Bayes mengasumsikan bahwa untuk setiap kata-kata yang terdapat pada suatu dokumen bersifat independen (tidak saling terkait) yang diberikan oleh kelas c, sehingga diperoleh persamaan [4]:

$$P(f_1, f_2, \dots, f_n|c) = P(f_1|c) \cdot P(f_2|c) \dots P(f_n|c) \quad (2.10)$$

Persamaan akhir klasifikasi Naive Bayes yang digunakan adalah:

$$(2.11) \quad c_{NB} = \operatorname{argmax}_{c \in C} P(c) \prod_{i=1}^n P(f_i|c)$$

Berdasarkan persamaan (2.11) langkah pertama yang dilakukan adalah menghitung probabilitas $P(c)$ dan juga nilai $P(f_i|c)$. Untuk menentukan $P(c)$ dapat dinyatakan dengan persamaan:

$$(2.12) \quad P(c) = \frac{N_c}{N_{doc}}$$

N_c adalah jumlah dokumen yang terdapat pada data latih untuk kelas c dan N_{doc} adalah jumlah seluruh dokumen pada data latih. Untuk mengetahui probabilitas $P(f_i|c)$ yang mengasumsikan kemunculan kata pada setiap dokumen dapat dinyatakan dengan persamaan [4] :

$$(2.13) \quad P(f_i|c) = \frac{\operatorname{count}(f_i,c)}{\sum_{f \in V} \operatorname{count}(f,c)}$$

V adalah seluruh kata-kata unik yang terdapat pada seluruh dokumen. Untuk menghindari terjadinya nilai nol pada probabilitas $P(f_i|c)$, maka dapat dinyatakan dengan persamaan [4] :

$$(2.14) \quad P(f_i|c) = \frac{\operatorname{count}(f_i,c) + 1}{(\sum_{f \in V} \operatorname{count}(f,c)) + V}$$

$\operatorname{count}(f_i,c)$ adalah jumlah kata tertentu yang terdapat pada seluruh dokumen pada kelas c , $\sum_{f \in V} \operatorname{count}(f,c)$ adalah jumlah seluruh kata yang terdapat pada kelas c [4] .

2.5 Confusion Matrix

Confusion matrix merupakan suatu metode yang digunakan untuk melakukan perhitungan akurasi pada konsep data mining. Metode ini mempunyai nilai keluaran yaitu recall, presisi, dan akurasi[4].

- a. *Recall* merupakan jumlah dokumen yang memang benar dan berhasil diprediksi secara benar. Recall digunakan sebagai alat ukur kuantitas item yang memang benar dan berhasil diprediksi dengan benar.

$$\operatorname{Recall} = TP / (FN + TP) \quad (2.15)$$

- b. Presisi merupakan jumlah dokumen yang diprediksi benar dan terbukti benar. Precision digunakan sebagai alat ukur ketepatan prediksi suatu sistem.

$$\operatorname{Precision} = TP / (FP + TP) \quad (2.16)$$

- c. Akurasi, adalah perbandingan kasus yang diidentifikasi benar dengan jumlah semua kasus.

$$\operatorname{Accuracy} = (TP + TN) / (TP + TN + FP + FN) \quad (2.17)$$

- d. *F-1 score* digunakan untuk mencari nilai *weighted average* antara *precision* dan *recall*. Perhitungan *F1-Score* dilakukan dengan rumus:

$$\operatorname{F1-Score} = \frac{2 \times \operatorname{Recall} \times \operatorname{Presisi}}{\operatorname{Recall} + \operatorname{Presisi}} \quad (2.18)$$

Keterangan:

TN : Data yang diklasifikasikan memiliki jumlah nilai yang diprediksi tidak benar dan faktanya juga menyimpan ketidakbenaran.

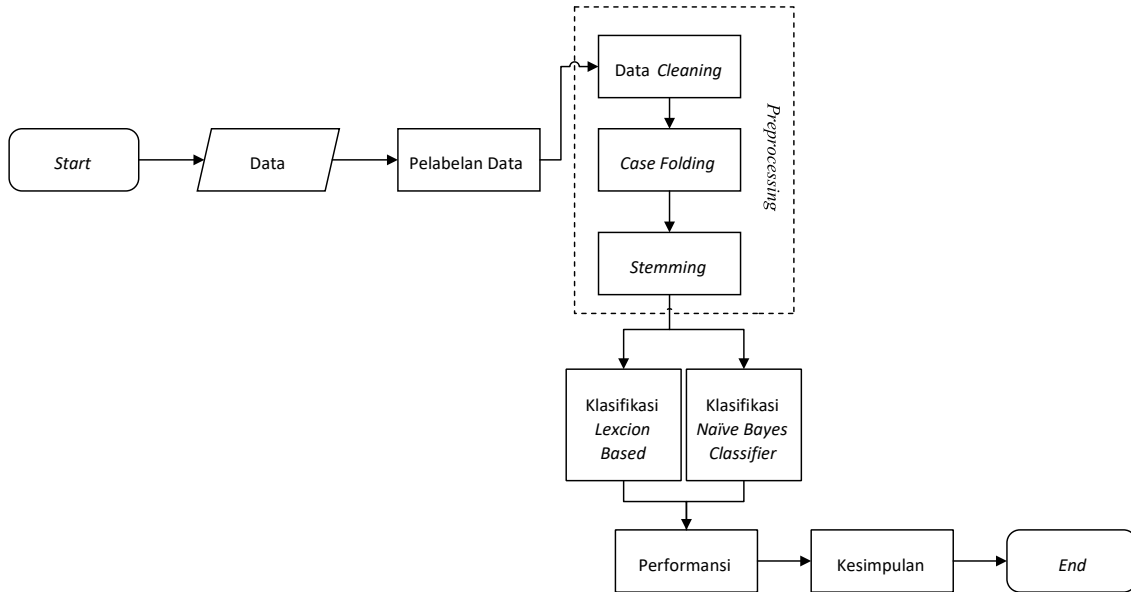
FP : Data yang diklasifikasikan memiliki jumlah nilai yang diprediksi benar namun faktanya berkata sebaliknya.

FN : Data yang diklasifikasikan memiliki jumlah nilai yang diprediksi tidak benar namun memiliki fakta yang benar

TP : Data yang diklasifikasikan memiliki jumlah nilai yang diprediksi benar dan faktanya ikut menyimpan kebenaran.

3. Sistem yang Dibangun

Secara umum sistem yang dibangun dalam penelitian ini adalah sebagai berikut.



Tabel 3.1 Metode Penelitian

3.1 Pengumpulan Data

Data set yang digunakan pada penelitian ini adalah komentar pada akun resmi media sosial Instagram tim @jokowi.amin dan @indonesiaadilmakmur. Data didapatkan dengan cara crawling memanfaatkan Instagram API (*application Programming Interface*) yang merupakan tool yang digunakan untuk mengakses pencarian tag, pencarian foto, feeds, dan komentar yang terdapat pada Instagram. Pengambilan data dilakukan mulai Januari sampai dengan Februari 2019 sebanyak 1434 data. Data yang didapatkan tidak seimbang sehingga harus dilakukan keseimbangan data setiap kelas. Jumlah data antar kelas harus dibuat seimbang (*balanced*) setiap kelasnya, karena dengan data yang tidak seimbang (*imbalanced*), klasifikasi yang dibangun memiliki kecenderungan untuk mengabaikan minority class [10]. Pada penelitian ini menggunakan under sampling untuk menyeimbangkan data tersebut. Sehingga, didapat data yang sudah seimbang sebanyak 360 data dan dibagi menjadi 80:20 untuk data training dan testing, yang terdiri dari 179 komentar *cyberbullying* dan 181 komentar *non cyberbullying*.

3.2 Pelabelan Data

Pelabelan data dilakukan untuk memberikan label pada komentar yang ditandai sebagai kalimat *non cyberbullying* ataupun *cyberbullying*. Untuk melabeli data harus memenuhi karakteristik yang telah ditentukan agar memudahkan partisipan dalam menentukan label.

- a. *Non cyberbullying* : Pada komentar tidak ada kata yang berhubungan dengan makna negatif ataupun ada komentar yang menggunakan kata kasar/ negatif tetapi tidak ditujukan kepada suatu objek.
- b. *Cyberbullying* : Terdapat berbagai komentar negatif yang ditunjukkan kepada suatu objek dengan tujuan mendiskriminasi seseorang, mengucilkan orang lain, serta mengungkapkan rasa kebencian kepada orang lain dengan menggunakan kata kasar.

Tabel 3.1 Contoh komentar & pelabelan data *non cyberbullying* dan *cyberbullying*.

Contoh Komentar	Responden 1		Responden 2		Responden 3	
	Cyberbullying	Non Cyberbullying	Cyberbullying	Non Cyberbullying	Cyberbullying	Non Cyberbullying
Bangga sama presidenku yg luar biasa bekerja demi kemajuan bangsa & negera..		Tidak		Tidak		Tidak
Siapa pilih jokowi		Tidak		Tidak		Tidak
Semangat pak... kerja saja. Warga yg mengkritik semoga jadi bahan evaluasi.		Tidak		Tidak		Tidak
pendukung prabowosandi guuwooblog	ya		ya		ya	
BANYAK BACOTT KAUU MANUSIA BAU TANAH	ya		ya		ya	
Banyak yang cabul pula anteknya Prabowo.	ya		ya		ya	

Pada penelitian ini menggunakan 3 orang mahasiswa sebagai partisipan yang akan melabeli komentar yang termasuk *cyberbullying* atau *non cyberbullying*. Partisipan melakukan pelabelan secara manual menggunakan google sheet. Pada penelitian ini digunakan 3 orang partisipan karena perbedaan pendapat antara masing-masing partisipan untuk menentukan komentar mana yang mengandung *cyberbullying* atau *non cyberbullying*. Data yang sudah diidentifikasi oleh partisipan dapat dilihat pada lampiran 1.

3.3 Preprocessing Data

Preprocessing merupakan suatu proses pengumpulan data mentah untuk diolah menjadi data yang bermanfaat. Pada data mentah perlu dilakukan proses data *preprocessing* karena masih mengandung data yang *incomplete* (data yang masih memiliki kekurangan pada atributnya), *noisy* (data yang masih mengandung *error* dan *outliers*) dan *inconsistent* (data yang tidak konsisten terhadap penamaan, singkatan, code)[9]. Proses *preprocessing* pada penelitian ini adalah sebagai berikut:

1. *Data Cleaning* dilakukan untuk menghapus karakter-karakter khusus, pembersihan kata dengan menghilangkan koma(,), titik(.), menghapus angka, menghapus URL, dan data-data kosong. Hal ini penting dilakukan karena proses pengambilan data otomatis dari Instagram tidak selalu berhasil dengan sempurna dan juga untuk mengurangi noise.
2. *Case folding*, yaitu mengkonversi keseluruhan teks menjadi bentuk standar, yaitu huruf kecil. Contohnya: KOMpuTer menjadi komputer.
3. *Stemming*, merupakan proses mencari kata dasar untuk memperkecil jumlah indeks yang berbeda dari suatu dokumen, dan juga untuk mengelompokkan kata yang memiliki kata dasar dan arti yang serupa. Contoh Stem (akar kata) adalah kata inti setelah imbuhan dihilangkan (awalan dan akhiran). Misalnya kata "perancangan" dan "merancang" akan diubah menjadi sebuah kata yang sama, yaitu "rancang". Proses *stemming* sangat tergantung kepada bahasa dari kata yang akan di-stem.

Tabel 3.2 Contoh Hasil Data *Cleaning*

Kalimat	Anjing kampungan Jokowi!!!
Hasil <i>Cleaning</i>	Anjing kampungan jokowi

Tabel 3.3 Contoh Hasil *Case folding*

Kalimat	Anjing kampungan jokowi
Hasil <i>Case folding</i>	anjing kampungan jokowi

Tabel 3.4 Contoh Hasil *Stemming*

Kalimat	anjing kampung jokowi
Hasil <i>Cleaning</i>	anjing kampung jokowi

3.4 Klasifikasi *Lexicon based*

Pada penelitian ini akan dilakukan identifikasi terhadap sejumlah kata yang terdapat pada komentar dimana diperlukan *sentiment lexicon* (kamus sentimen) sebagai acuan dalam menentukan sentimen yang terkandung pada komentar serta polaritas masing-masing kata, sehingga dapat terklasifikasi apakah termasuk sentimen positif atau sentimen negative. Polaritas suatu kata yang di dapat akan digunakan sebagai acuan untuk melakukan proses perbandingan, sehingga dalam satu kalimat akan diketahui total jumlah nilai positif dan juga nilai negative dari tiap-tiap kata penyusunnya. Jika total jumlah nilai positif lebih besar dari jumlah nilai negatif maka kalimat akan berorientasi positif. Namun jika total jumlah nilai positif kurang dari jumlah nilai negatif maka kalimat akan berorientasi negatif. Penelitian ini menggunakan kamus yang sudah ada yaitu *sentiment lexicon* berbahasa Indonesia yang dikembangkan oleh [11] [12]. Contoh dibawah ini untuk menentukan suatu kalimat bersentimen positif atau negatif berdasarkan kamus *lexicon*, Ada beberapa tahapan yang dilakukan. Pada penelitian ini tahapan penentuan sentimen yakni, penentuan polaritas kata, penanganan negasi, serta pemberian skor terhadap kalimat.

Menentukan kata untuk klasifikasi dengan *Lexicon Based*:

1. Setiap kata dalam kalimat akan diberi sebuah nilai yakni bernilai sesuai dengan polaritas yang sudah ditetapkan di kamus untuk kata positif dan negatif. Kemudian didapatkan total kata bersentimen pada kalimat tersebut.
2. Penanganan kata negasi: kata negasi seperti kata “tidak” pada kalimat contoh “Anda tidak bahagia” akan membalikkan orientasi sentimen. Dengan kata lain, kata bersentimen positif “bahagia” akan diberi nilai negative.
3. Pemberian skor pada kalimat: Skor digunakan untuk menentukan apakah sebuah kalimat bersentimen positif atau bersentimen negative.

Tabel 3.5 Contoh *Lexicon Based*

Kalimat	Kata pada kamus positif	Kata pada kamus negatif	Hasil
Anjing kampung jokowi	Jokowi (0)	Anjing (-3), kampung (-3)	-6

Didapatkan hasil negative untuk kalimat diatas, sehingga tergolong dalam kelas *cyberbullying*.

3.5 Klasifikasi *Naïve Bayes Classifier*

3.5.1 Pengujian Data Latih *Naïve Bayes Classifier*

Untuk mengetahui nilai probabilitas, data yang digunakan yaitu data latih yang telah mempunyai kelas aktual serta telah melewati tahap *preprocessing* terlebih dahulu. Data latih yang digunakan yaitu sebanyak 3 komentar dengan 1 kelas *cyberbullying* dan 2 kelas non *cyberbullying*. Nilai probabilitas dapat dihitung menggunakan persamaan (2.4):

$$P(c) = \frac{Nc}{Ndoc}$$

$$P(c = cyberbullying) = \frac{1}{3} = 0.333$$

$$P(c = non cyberbullying) = \frac{2}{3} = 0.666$$

Setelah nilai probabilitas diperoleh, selanjutnya yaitu menghitung nilai probabilitas $P(f_i|c)$ yang ditentukan dengan persamaan (2.14) dengan banyak kata istilah yang ditentukan adalah 6 kata. berikut beberapa hasil perhitungan ditunjukkan:

Sehingga diperoleh:

$$P(f_i|c) = \frac{\text{count}(f_i,c) + 1}{(\sum_{f \in V} \text{count}(f,c)) + V}$$

$$P(f_i = \text{dasar}|c=\text{cyberbullying}) = \frac{1+1}{3+|6|} = 0.222$$

$$P(f_i = \text{kampung}|c=\text{cyberbullying}) = \frac{1+1}{3+|6|} = 0.222$$

$$P(f_i = \text{jokowi}|c=\text{cyberbullying}) = \frac{1+1}{3+|6|} = 0.222$$

$$P(f_i = \text{presiden}|c=\text{noncyberbullying}) = \frac{2+1}{5+|6|} = 0.272$$

$$P(f_i = \text{dua}|c=\text{noncyberbullying}) = \frac{1+1}{5+|6|} = 0.181$$

$$P(f_i = \text{periode}|c=\text{noncyberbullying}) = \frac{1+1}{5+|6|} = 0.181$$

$$P(f_i = \text{jokowi}|c=\text{noncyberbullying}) = \frac{1+1}{5+|6|} = 0.181$$

$$P(f_i = \text{mantap}|c=\text{noncyberbullying}) = \frac{1+1}{5+|6|} = 0.181$$

3.5.2 Pengujian Data Uji Naive Bayes Classifier

Berikut ini adalah hasil kelas prediksi menggunakan *Naive Bayes Classifier* berdasarkan nilai probabilitas $P(c)$ dan $P(f_i|c)$ pada data latih. Sehingga diperoleh nilai cNB pada data uji. cNB ini diuji pada setiap komentar dengan diuji pada masing-masing kelas *cyberbullying* dan *noncyberbullying*. Apabila hasil cNB *cyberbullying* lebih besar daripada cNB *noncyberbullying*, maka komentar tersebut tergolong kelas *cyberbullying*. Apabila cNB positif lebih kecil daripada cNB *noncyberbullying*, maka komentar tersebut tergolong kelas *noncyberbullying*.

contoh kata-kata seperti kambing, kampung, bacot, dan anjing. Untuk mengetahui kelas dari komentar tersebut dibutuhkan nilai $P(w|c)$ dari hasil data latih. Pada komentar tersebut terdapat kata anjing dan kampung sebagai kata yang sudah memiliki nilai probabilitas:

Perhitungan kelas *cyberbullying* dengan kata ‘anjing kampung jokowi’

$$P(f_i = \text{anjing}|c = \text{cyberbullying}) = 0,111$$

$$P(f_i = \text{kampung}|c = \text{cyberbullying}) = 0.0090$$

$$P(f_i = \text{jokowi}|c = \text{cyberbullying}) = 0.352$$

$$P(c = \text{cyberbullying}) = 0,7569$$

$$\text{cNB } \text{cyberbullying} = 0.111 * 0.0090 * 0.352 * 0.7569 = 0.000266$$

Perhitungan kelas *noncyberbullying*

$$P(f_i = \text{anjing} | c = \text{noncyberbullying}) = 0.0007$$

$$P(f_i = \text{kampung} | c = \text{noncyberbullying}) = 0.0040$$

$$P(f_i = \text{jokowi} | c = \text{noncyberbullying}) = 0.1140$$

$$P(c = \text{noncyberbullying}) = 0.2410$$

$$cNB_{\text{noncyberbullying}} = 0.0007 * 0.0040 * 0.1140 * 0.2410 = 0.00000008$$

Berdasarkan hasil perhitungan diatas karena cNB positif lebih besar daripada cNB negatif maka hasil komentar tersebut termasuk ke dalam kelas *cyberbullying*.

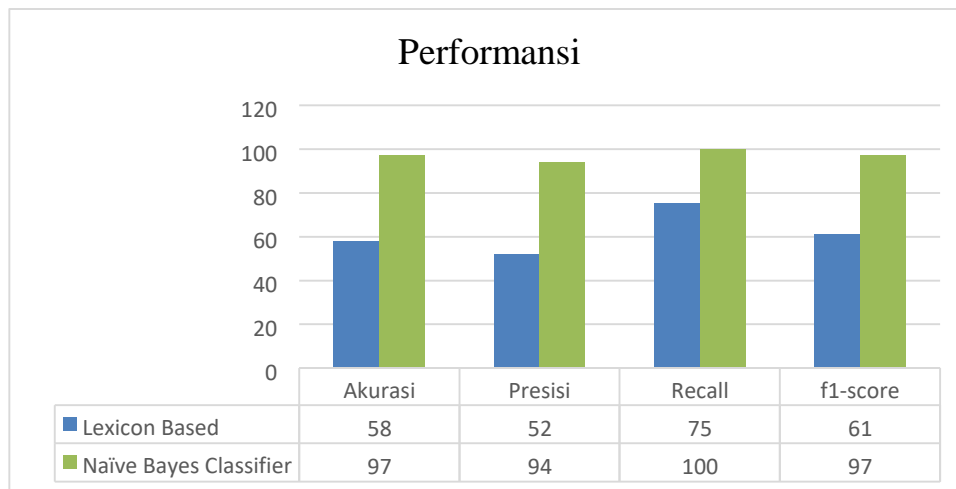
4. Evaluasi

4.1 Skenario Pengujian

Pengukuran kinerja dari sistem / metode klasifikasi dilakukan dengan menghitung nilai akurasi, *recall*, presisi, dan *F1-score* pada kedua metode. Akurasi diperoleh dengan membandingkan jumlah data hasil klasifikasi (prediksi) yang sesuai dengan jumlah keseluruhan data. Semakin tinggi nilai akurasi yang diperoleh, maka hasil klasifikasi semakin baik. Akan tetapi, jika hanya melihat nilai akurasi saja tidak bisa mendeteksi adanya penyimpangan data. Oleh karena itu penelitian ini juga menghitung nilai *recall* dan presisi. *Recall* diperoleh dengan membandingkan jumlah data hasil klasifikasi yang relevan dan total data yang dianggap relevan. Presisi diperoleh dengan membandingkan jumlah data hasil klasifikasi yang relevan dan total jumlah data yang ditemukan pada kelas tertentu. Sedangkan *F1-score* digunakan untuk mengetahui keseimbangan antara presisi dan *recall* yang didapat dari sistem yang akan dibangun. Perhitungan yang digunakan untuk menghitung nilai akurasi, *recall*, dan presisi ditunjukkan pada persamaan 2.15 sampai dengan 2.18. Hal ini dikarenakan pada penelitian ini hanya mengidentifikasi komentar yang mengandung *cyberbullying* dan *non cyberbullying*, sehingga identifikasi dibutuhkan untuk mengetahui apakah hasil prediksi sesuai dengan sistem. Pada pengujian ini, perbandingan jumlah data yang digunakan untuk metode *Naïve Bayes Classifier* adalah 80:20 dimana dari total 360 data, digunakan 288 data *training* yang sebelumnya sudah diketahui kategorinya dan 72 data *testing* yang digunakan untuk prediksi klasifikasi atau data yang belum diketahui klasifikasinya.

4.2 Pengukuran Performansi

Hasil analisis dan pengujian yang telah dilakukan menggunakan data latih sebanyak 360 data dari Identifikasi *Cyberbullying* pada Komentar Instagram menggunakan Metode *Lexicon Based* dan *Naïve Bayes Classifier* dengan studi kasus Pemilihan Presiden Indonesia Tahun 2019 menghasilkan nilai akurasi 58% presisi 52%, *recall* 75% dan *F-1 Score* sebesar 61%. Sedangkan pengujian menggunakan metode *Naïve Bayes Classifier*, memberikan nilai persentase akurasi, presisi, *recall*, dan *F1-score* lebih tinggi daripada *Lexicon Based*, karena metode *Naïve Bayes Classifier* dalam melakukan pengujian klasifikasi, memiliki tingkat ketepatan yang tinggi meskipun sangat sederhana dan efisien. Selain itu, hasil dari ketepatan klasifikasi menggunakan metode *Naïve Bayes Classifier* bergantung pada jumlah data latih yang digunakan dalam sistem. Hasil analisis dari pengujian data menggunakan metode *Naïve Bayes Classifier*, diperoleh nilai presisi terbesar yaitu 97%, *recall* 100%, akurasi 97% dan *F1-score* 97%. Berdasarkan analisa diatas, dapat disimpulkan bahwa kinerja *Naïve Bayes Classifier* dalam melakukan klasifikasi terbukti lebih baik daripada menggunakan metode klasifikasi *Lexicon Based*. Hasil tersebut dapat dilihat pada Tabel 4.1.



Tabel 4.1 Grafik performansi *Lexicon-Based* dan *Naïve Bayes Classifier*

Pada Tabel 4.1 dapat dilihat bahwa nilai akurasi keseluruhan yang menunjukkan tingkat kedekatan antara nilai prediksi dengan hasil aktual adalah 97% untuk metode *Naïve Bayes Classifier*, yang menunjukkan persentase semua prediksi yang benar dari keseluruhan data. Kemudian nilai presisi menggunakan metode *Naïve Bayes Classifier* menunjukkan nilai persentase yang lebih besar daripada nilai presisi menggunakan metode *Lexicon Based* yaitu 94%, hal ini disebabkan nilai tingkat ketepatan data actual dengan hasil yang diberikan oleh sistem menunjukkan nilai yang lebih besar. Nilai persentase terbesar dari recall yaitu 100% dengan metode klasifikasi yang digunakan yaitu *Naïve Bayes Classifier*, besarnya nilai persentase *recall* disebabkan oleh tingkat keberhasilan sistem dalam menemukan data.

Pada *Lexicon Based* menunjukkan bahwa bahwa hasil pengujian tidak lebih baik dibandingkan metode *naïve bayes classifier*, hal ini bisa disebabkan karena ada data *cyberbullying* yang tergolong sebagai *noncyberbullying* maupun sebaliknya. Contohnya pada kalimat ‘kubu cebong kok gblg’ pada kata cebong yang arti sebenarnya adalah tahap pra-dewasa dalam daur hidup amfibia. Tetapi dalam kasus pilpres ini cebong merupakan julukan terhadap pendukung jokowi. Kemudian penggunaan kata yang tidak baku yang tidak terdeteksi oleh sistem. walaupun pada penelitian ini kamus bahasa gaul sudah diterapkan. Sehingga keterbatasan kamus yang digunakan sebagai identifier bisa mempengaruhi hasil klasifikasi yang kurang baik.

5. Kesimpulan

Berdasarkan penelitian yang telah dilakukan, maka diperoleh kesimpulan sebagai berikut:

1. Metode *Lexicon based* dan *Naïve Bayes Classifier* dapat digunakan untuk pengklasifikasian komentar *cyberbullying* dan *non cyberbullying* terutama pada komentar yang membahas tentang Pilpres 2019.
2. Hasil pengujian pada sistem yang dibangun memperlihatkan bahwa metode *naïve bayes classifier* memberikan akurasi, presisi, *recall*, dan *f1-score* yang lebih baik daripada metode *lexicon based*. Dengan nilai akurasi 97%, presisi 94%, *recall* 100% dan *F1-score* 97%.
3. Pada *lexicon-based* tidak diperlukan training terhadap data sehingga sangat bergantung kepada kamus. Jika kamus lengkap maka performa sistem akan baik, sebaliknya jika kamus tidak lengkap maka performa sistem akan buruk. Dari pengujian yang dilakukan terhadap sistem, menghasilkan tingkat akurasi sebesar 58%, ini berarti kamus belum cukup mewakili kata kunci-kata kunci yang ada dalam kalimat pada kasus identifikasi *cyberbullying* untuk kasus pemilihan presiden Indonesia tahun 2019.

6. Saran

Adapun saran yang dipertimbangkan untuk pengembangan penelitian selanjutnya sebagai berikut:

1. Penambahan kata pada kamus sentiment sesuai konteks masalah yang akan diidentifikasi sehingga akan meningkatkan perolehan akurasi.
2. Penelitian-penelitian selanjutnya dapat dilakukan pembaharuan kamus secara otomatis dengan cara mengkombinasikan *lexicon-based* dengan *machine learning*.

Daftar Pustaka

- [1] Asosiasi Penyelenggara Jasa Internet Indonesia, “Survei APJII: Penetrasi Internet di Indonesia Capai 143 Juta Jiwa,” *Apjii.or.Id.* p. 3, 2018.
- [2] T. H. E. Annual and B. Of, “Annual Bullying,” 2018.
- [3] Undang-Undang, “Uu-2008-11 Informasi Dan Transaksi Elektronik,” *Undang-undang*, vol. 11, pp. 1–18, 2008.
- [4] R. Afrinza and K. M. Lhaksmana, “DETEKSI CYBER BULLYING PADA TWITTER,” pp. 4–14.
- [5] Y. S. Mahardhika, E. Zuliarso, P. Studi, T. Informatika, F. T. Informasi, and U. Stikubank, “ANALISIS SENTIMEN TERHADAP PEMERINTAHAN JOKO WIDODO PADA MEDIA SOSIAL TWITTER MENGGUNAKAN ALGORITMA NAIVES BAYES,” no. 2015, pp. 409–413, 2018.
- [6] M. Z. Naf’an, A. A. Bimantara, A. Larasati, E. M. Risondang, and N. A. S. Nugraha, “Sentiment Analysis of Cyberbullying on Instagram User Comments,” *J. Data Sci. Its Appl.*, vol. 2, no. 1, pp. 88–98, 2019.
- [7] W. A. Luqyana, I. Cholissodin, and R. S. Perdana, “Analisis Sentimen Cyberbullying Pada Komentar Instagram dengan Metode Klasifikasi Support Vector Machine,” *J. Pengemb. Teknol. Inf. dan Ilmu Komput. Univ. Brawijaya*, vol. 2, no. 11, pp. 4704–4713, 2018.
- [8] Peng, W., & Park, D.H. (2011). Generate Adjective Sentiment Dictionary for Social Media Sentiment Analysis Using Constrained Nonnegative Matrix Factorization. *ICWSM*.
- [9] A. H. Primasanti, R. Agustiansyah, and A. Fitrianyah, “PEMBOBOTAN DATA PREPROCESSING DALAM ANALISIS SENTIMEN MENGGUNAKAN METODE LEXICON-BASED (Studi Kasus : Pilkada Jawa Timur 2018) WEIGHTING OF DATA PREPROCESSING IN SENTIMENT ANALYSIS USING LEXICON-BASED METHOD (Case Study : The 2018 East Java Regional He,” 2018.
- [10] F. Rayhan *et al.*, “MEBoost: Mixing estimators with boosting for imbalanced data classification,” *Int. Conf. Software, Knowl. Information, Ind. Manag. Appl. Ski.*, vol. 2017-Decem, no. December, 2018.
- [11] B. Liu, S. M. Street, S. M. Street, and S. M. Street, “WWW05-Opinion Observer Analyzing and Comparing Opinions.pdf.”
- [12] E. T. Iede, S. R. C. Penteado, and J. C. Bisol, “Primeiro registro de ataque de *Sirex noctilio* em *Pinus taeda* no Brasil,” *Circ. Técnica EMBRAPA-CNPF n. 20. EMBRAPA, Brazil*, vol. 10, no. 2, pp. 207–218, 1988.

