

Analisis *Parsing* Data Sosial Media Terhadap Konflik Kelompok Kriminal Bersenjata Di Papua Dengan Pendekatan *Lexicon Based*

1st Adiev Ghifari Syahviar
Fakultas Ilmu Terapan
Universitas Telkom
Bandung, Indonesia

ghifarisyah@student.telkomuniversity.ac.id

2nd Setia Juli Irzal Ismail
Fakultas Ilmu Terapan
Universitas Telkom
Bandung, Indonesia

julismail@telkomuniversity.ac.id

3rd Muhammad Rizqy Alfarisi
Fakultas Ilmu Terapan
Universitas Telkom
Bandung, Indonesia

mrizkyalfarisi@telkomuniversity.ac.id

Abstrak - Aksi kriminal dan separatistis yang telah dilakukan oleh Kelompok Kriminal Bersenjata di Papua sejak tahun 1965 selalu menimbulkan korban. Pada 2 maret 2022 telah terjadi penembakan terhadap 8 karyawan PT Palaparing Timur Telematics (PTT). Dari peristiwa tersebut menuai berbagai opini masyarakat, yang terdiri dari opini positif hingga opini negatif. Untuk mengetahui opini masyarakat tentang peristiwa tersebut, dilakukan analisis parsing data dari sosial media. Analisis parsing data pada penelitian ini menggunakan 1.821 data tweet yang diambil pada bulan Agustus 2022 melalui platform media sosial twitter sebagai sumber data sentimennya. Pada penelitian ini dilakukan tahapan proses seperti pengambilan data Twitter, pengolahan data, klasifikasi data dan Evaluasi model. Untuk klasifikasi data teks pada penelitian ini menggunakan pendekatan berbasis leksikon yaitu dengan menggunakan library VADER dan kamus leksikon manual yang disusun oleh Professor Bing Liu dari Universitas Chicago. Hasil analisis berdasarkan VADER sentimen menghasilkan informasi bahwa sebagian besar 47,1% masyarakat Indonesia memberikan tanggapan positif terhadap konflik kelompok kriminal bersenjata di Papua dan mendapatkan hasil akurasi dari proses evaluasi data secara keseluruhan sebesar 80,7 %.

Kata Kunci - kelompok kriminal bersenjata, papua, analisis parsing data, vader sentimen, leksikon

I. PENDAHULUAN

Kemajuan teknologi informasi dan komunikasi saat ini sudah berkembang sangat pesat, terutama perangkat digital dan sosial media yang sudah menjadi kebutuhan primer bagi masyarakat. Salah satu contoh dari kemajuan teknologi informasi dan komunikasi pada saat ini adalah dengan adanya platform sosial media seperti Facebook, Twitter, Instagram, Telegram, Tiktok dan YouTube. Dari platform sosial media tersebut, salah satu platform sosial media yang paling cepat pertumbuhannya adalah twitter [1]. Twitter merupakan salah satu media sosial yang sering

digunakan oleh masyarakat untuk membagikan informasi, cerita, kegiatan dan emosi-emosi yang mereka rasakan [2].

Twitter masih ramai diminati sampai saat ini dikarenakan mempunyai kelebihan dari platform media sosial lainnya berupa fitur-fitur menarik, seperti memudahkan untuk membisukan kata-kata berupa cuitan *Toxic* atau kata negatif, lalu twitter mempunyai fitur trending topic. Berdasarkan data dari PT Bakrie Telecom, Indonesia memiliki 19,5 juta pengguna twitter di Indonesia dari total 500 juta pengguna global [3]. Kemudian Selamatta Sembiring selaku Direktur Pelayanan Informasi Internasional Ditjen Informasi dan Komunikasi Publik (IKP), beliau mengatakan bahwa situs jejaring sosial yang paling banyak diakses oleh masyarakat Indonesia adalah Facebook dan Twitter [3]. Tidak heran jika Indonesia saat ini menempati peringkat 5 pengguna Twitter terbesar di dunia, setelah USA, Brazil, Jepang dan Inggris [3].

Selain untuk alat berkomunikasi antar sesama pengguna, twitter dapat dimanfaatkan juga sebagai wadah atau media untuk melakukan penelitian analisis sentimen. Analisis sentimen memiliki tujuan untuk dapat memahami bagaimana orang menggambarkan dan mengungkapkan persepsi mereka terhadap suatu kasus baik secara positif, negatif maupun netral [4]. Melalui twitter dengan hashtag #KKBPAPUA terdapat cuitan yang berisi berita duka yaitu pembunuhan terhadap 8 karyawan Palaparing Timur Telematika (PTT) oleh Kelompok Kriminal Bersenjata pada hari Rabu, tanggal 2 Maret 2022 di Tower B3, Kampung Jenggeran, Distrik Beoga, Kabupaten Puncak. Dari peristiwa mengerikan tersebut, satu orang berhasil melarikan diri dan sudah di evakuasi. Kasus pembunuhan tersebut dilakukan oleh kelompok separatistis yang dikenal dengan istilah KKB (Kelompok Kriminal Bersenjata), Sudah bukan hal baru bagi mereka melakukan aksi-aksi meresahkan tersebut, karena sejak awal terbentuknya KKB (Kelompok Kriminal Bersenjata) pada tahun 1965, mereka

sudah sering melakukan pergerakan, salah satu pergerakannya yaitu dengan metode gerilya dalam upaya pemberontakan [5]. Banyak masyarakat Indonesia yang memberikan tanggapan mengenai pembahasan terkait konflik yang terjadi di Papua, berbagai macam respon dalam bentuk kicauan *tweet* yang tentunya mengandung sentimen positif ataupun negatif.

Berdasarkan latar belakang diatas maka dibuatlah sebuah penelitian yang berjudul "ANALISIS PARSING DATA SOSIAL MEDIA TERHADAP KONFLIK KELOMPOK KRIMINAL BERSENJATA DI PAPUA DENGAN PENDEKATAN *LEXICON BASED*". Penelitian ini menggunakan pendekatan leksikon berbasis *VADER* untuk menganalisis polaritas opini masyarakat. Manfaat dari penelitian analisis *parsing* data sosial media yang dibuat ini sangat penting sebagai alat bantu untuk mencari kata sentimen negatif dan positif dari opini masyarakat Indonesia yang berkaitan dengan konflik kelompok kriminal bersenjata di Papua.

A. Rumusan Masalah

Bagaimana mencari kata sentimen negatif dan positif yang berkaitan dengan konflik kelompok kriminal bersenjata di Papua pada sosial media dan bagaimana mengukur akurasi hasil analisa kata sentimen yang dilakukan. Untuk itu dibutuhkan suatu tools dan metode dalam menganalisa opini masyarakat Indonesia terhadap konflik kelompok kriminal bersenjata di Papua.

B. Tujuan

Berdasarkan rumusan masalah, maka tujuan dari proyek akhir ini sebagai berikut:

1. Mendapatkan hasil klasifikasi opini masyarakat Indonesia terhadap konflik kelompok kriminal bersenjata di Papua melalui sosial media twitter dengan cara *memparsing* data tweet kemudian dilakukan klasifikasi menggunakan pendekatan *lexicon based* berbasis *VADER*.
2. Mengukur akurasi hasil analisa sentimen.

C. Batasan Masalah

Adapun yang menjadi batasan dari permasalahan pada proyek akhir ini adalah sebagai berikut:

1. Pengumpulan dan pengambilan data hanya dari twitter.
2. Tidak adanya pengecekan terhadap profil pengguna.
3. Emoji dan tanda baca yang terkandung disetiap data tweet tidak digunakan untuk penelitian analisis sentimen ini.
4. Tidak melakukan pengecekan dan pembahasan mengenai akun buzzer.
5. Data tweet yang digunakan untuk pelabelan *VADER* dan data tweet yang ditampilkan pada hasil akhir penelitian diterjemahkan bahasa Inggris.

6. Penggunaan kamus leksikon pada penelitian ini menggunakan kamus leksikon yang sudah tersedia.
7. Sentimen positif dari sebagian opini masyarakat Indonesia terhadap konflik kriminal bersenjata di Papua adalah dengan asumsi mendukung gerakan Papua untuk merdeka.
8. Sentimen netral dari sebagian opini masyarakat Indonesia pada penelitian ini didapat berdasarkan kata-kata yang tidak terdeteksi atau tidak terdapat pada kamus manual dan kamus *VADER*.
9. Sentimen negatif pada penelitian ini didapat berdasarkan banyaknya katakata negatif pada suatu kalimat berdasarkan kamus manual dan kamus *VADER*.

II. KAJIAN TEORI

A. Parsing

Parsing atau bisa juga disebut analisa sintaksis merupakan suatu proses penentuan struktur sebuah kalimat berdasarkan *grammar* dan *lexicon* tertentu [28].

B. Vader (*Valance Aware Dictionary and sEntiment Reasoner*)

Valance Aware Dictionary and sEntiment Reasoner atau disingkat *Vader* merupakan alat analisis sentimen yang berbasis aturan secara khusus yang disesuaikan dengan sentimen yang sudah diekspresikan di media sosial [14]. Kemampuan klasifikasi dilakukan *VADER* dengan cara memberikan nilai pada setiap kata dalam teks yang nantinya dapat menghasilkan output positif, negatif, netral [15]. Kelebihan pada metode *VADER* adalah mampu menangkap makna tersirat dari tanda baca yang terdapat dalam sebuah teks [15].

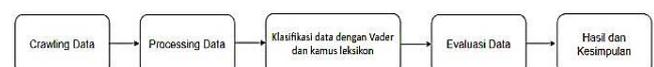
C. Lexicon-Based

Metode dengan pendekatan yang sangat bergantung pada kamus leksikon sentimen, yaitu suatu kumpulan kata sentimen yang diketahui dan dikompilasi sebelumnya. [7] [20]

D. Confusion Matrix

Confusion Matrix adalah sebuah metode yang biasa digunakan untuk perhitungan akurasi, *recall*, *precision*, dan *error rate*. *Accuracy* merupakan perbandingan kasus yang diidentifikasi benar dengan jumlah seluruh kasus dan *error rate* merupakan kasus yang diidentifikasi salah dengan umlah seluruh kasus [24].

III. METODE

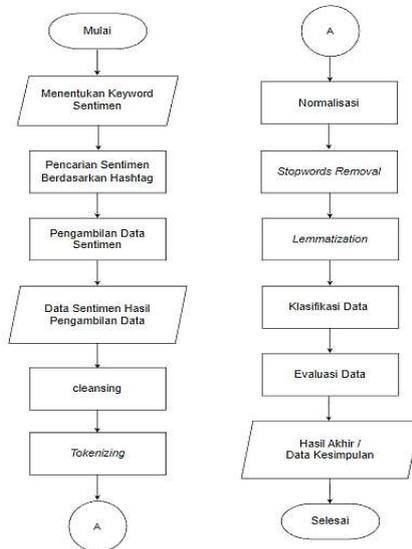


GAMBAR 1
(RANCANGAN SISTEM YANG AKAN DIBANGUN)

A. Gambar Sistem Usulan

Rancangan sistem yang akan dibangun pada gambar 1 merupakan rancangan sistem secara keseluruhan dari penelitian yang akan dibuat. Pada rancangan sistem analisis *parsing data* yang dibuat akan menerapkan *library Vader* untuk proses klasifikasi data yang nantinya akan menghasilkan *output* yang terbagi ke dalam tiga *class sentiment* seperti positif, negatif, dan netral.

B. Flowchart Sistem



GAMBAR 2
(FLOWCHART SISTEM)

Flowchart pada gambar 2 menjelaskan bahwa sistem dimulai dengan melakukan input untuk menentukan sebuah *keyword* sentimen yang dicari berdasarkan hashtag twitter. Pengambilan data yang telah dilakukan akan di proses secara berurutan yang dimulai dari tahap *cleansing*, lalu penerapan *case folding* untuk melakukan penyamarataan seluruh huruf kapital menjadi huruf kecil. Selanjutnya dilakukan *tokenizing* yaitu pemecahan kalimat menjadi kata, kemudian dilanjutkan ke tahap normalisasi untuk memperbaiki teks dengan penulisan yang keliru, lalu *stopwords removal* yaitu menghapus kata-kata imbuhan dan kata sambung. Dilanjutkan dengan proses menerjemahkan data tweet tersebut ke dalam bahasa inggris menggunakan fitur *google translate* yang tersedia pada *google spreadsheets*, lalu pada tahap *text preprocessing* yang terakhir yaitu *lemmatization* yang berfungsi untuk mengubah suatu kata menjadi kata dasar, setelah itu akan dilakukannya klasifikasi data menggunakan *library VADER*.

C. Tahapan Pengerjaan

Penelitian ini menggunakan *library Vader (Valance Aware Dictionary and sEntiment Reasoner)* untuk melakukan proses pelabelan dan mengklasifikasikan data *tweet*. Pada Gambar 3 merupakan tahapan pengerjaan dari sistem analisis *parsing data*, adapun penjelasan dari masing-masing tahapan pengerjaan mengenai sistem analisis *parsing data* secara keseluruhan sebagai berikut:

1. Studi literatur pada penelitian sebelumnya.

Melakukan pemahaman lebih lanjut terkait fokus masalah yang akan diteliti.

2. *Crawling Data Tweet*

Melakukan pengambilan dan pengumpulan data menggunakan *twint*, akan tetapi sebelum melakukan pengambilan data harus terlebih dahulu menentukan kata kunci untuk memperoleh data-data *tweet* yang didapat melalui *tools twint*. Kata kunci diperoleh melalui hashtag dengan menggunakan metode pengumpulan *tweets* manual yaitu memasukkan berbagai macam kata kunci yang berhubungan dengan konflik KKB di Papua.

1. Pengolahan data

Adapun tahap-tahap pengolahan data yaitu: *cleansing*, *case folding*, *tokenizing*, normalisasi, *stopwords removal*, dan *lemmatization*.

2. Klasifikasi data

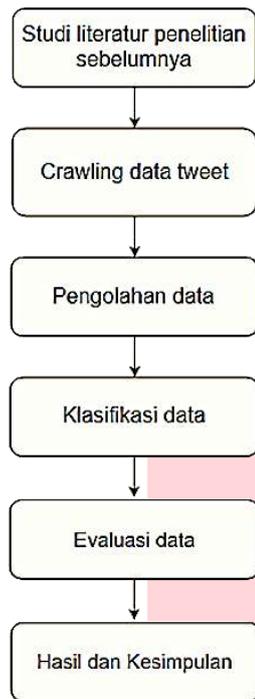
Data yang telah selesai diolah akan diklasifikasikan menggunakan metode *Lexicon-Based* menggunakan *library VADER (Valance Aware Dictionary and sEntiment Reasoner)* yang akan terbagi menjadi 3 kelas sentimen yaitu positif, negatif, dan netral. Kemudian akan dilakukan klasifikasi pelabelan manual menggunakan kamus leksikon dengan cara mencocokkan hasil *Vader Sentiment* dengan kata yang ada pada kamus leksikon manual, dari hasil klasifikasi tersebut akan dilakukannya evaluasi data untuk mendapatkan hasil akhir penelitian.

3. Evaluasi data

Evaluasi data pada penelitian ini menggunakan *confusion matrix*, *confusion matrix* ini biasanya digunakan pada proses klasifikasi data sebagai metode evaluasi visual. Dari tahap evaluasi data ini nantinya akan menghasilkan nilai akurasi dari proses perhitungan tabel *confusion matrix*.

4. Hasil dan Kesimpulan

Setelah evaluasi data selesai, selanjutnya akan dilakukan pengambilan kesimpulan untuk hasil akhir dari penelitian ini.



GAMBAR 3 (TAHAPAN Pengerjaan)

Kebutuhan Perangkat Lunak dan Keras

5. Perangkat Lunak

Software atau perangkat lunak yang dibutuhkan dalam membangun rancangan Proyek Akhir ini diantaranya sebagai berikut.

TABEL 1 (PERANGKAT LUNAK)

| NO | SOFTWARE | FUNGSI |
|----|--------------------|---|
| 1 | Anaconda Navigator | Menjalankan aplikasi Jupyterlab |
| 2 | Jupyterlab | Membuat program analisis sentimen |
| 3 | Twint | Sebagai media untuk melakukan pengumpulan data twitter |
| 4 | Twitter | Sumber informasi untuk pengambilan data-data sentimen |
| 5 | Python | Bahasa pemrograman yang digunakan untuk membuat program analisis sentimen |

6. Perangkat Keras

Perangkat keras yang dibutuhkan untuk membangun sistem ini ialah sebagai berikut.

TABEL 2 (PERANGKAT KERAS)

| NO | HARDWARE | FUNGSI |
|----|-----------|--|
| 1 | Laptop/PC | Menjalankan program analisis parsing data dan menjalankan aplikasi-aplikasi yang dibutuhkan. |

IV. HASIL DAN PEMBAHASAN

A. Klasifikasi Kamus Leksikon Manual

Setelah mendapatkan hasil dari klasifikasi data tweet menggunakan VADER, hasil tersebut digunakan untuk melakukan klasifikasi menggunakan kamus leksikon manual. Untuk dataset kamus lexicon manual yang digunakan pada penelitian ini menggunakan kamus buatan dari Professor Bing Liu [23]. Pada pelabelan VADER hanya bisa melakukan pelabelan kata sentimen dengan bahasa Inggris saja, maka kamus lexicon yang digunakan pada penelitian ini juga harus berbahasa Inggris. Oleh karena itu pada tahap sebelumnya data tweet diterjemahkan terlebih dahulu ke dalam Bahasa Inggris menggunakan fitur yang terdapat pada google spreadsheet.

Sebelum klasifikasi akan dilakukan, diperlukan pengelompokan data tweet sentimen sesuai dengan kelas sentimennya masing-masing berdasarkan hasil dari klasifikasi VADER, agar data lebih mudah untuk di proses ke tahap berikutnya. Klasifikasi menggunakan python library Vader sentiment default dengan batas skor polaritas positif lebih besar dari 0,5 dan negatif lebih besar dari -0,5 menghasilkan sentimen positif sebanyak 780, sentimen negatif sebanyak 198 dan sentimen netral sebanyak 843. Lalu dilakukan percobaan klasifikasi Vader Custom dengan batas skor polaritas positif 0,4 dan negatifnya -0,3 menghasilkan sentimen positif sebanyak 859, sentimen negatifnya sebanyak 297 dan sentimen netral sebanyak 665. Kemudian hasil positif menggunakan metode lexicon-based sebanyak 887, Negatif sebanyak 347, dan Netral sebanyak 587. Ini menunjukkan terdapat perbedaan dalam mengklasifikasikan kalimat seperti data tweet pada twitter. Hal ini terjadi karena klasifikasi dengan pendekatan lexicon-based bekerja dengan cara mencocokkan kata pada setiap kalimatnya dengan kata yang ada pada kamus lexicon.

Kemudian klasifikasi akan dilanjutkan dengan membandingkan nilai positif dan negatif dari kalimat tersebut. Jika jumlah kata positif lebih banyak dari jumlah kata negatif maka kalimat tersebut akan diklasifikasikan ke dalam kelas sentiment positif. Begitu juga sebaliknya, jika kata negatif lebih banyak dari kata positif pada suatu kalimat maka akan diklasifikasikan ke dalam kelas sentiment negatif. Kelas sentimen netral diklasifikasikan apabila jumlah kata positif dan negatif pada suatu kalimat berjumlah sama banyak atau bisa juga kasusnya tidak adanya kata positif dan negatif di dalam kalimat tersebut.

Berikut pada gambar 4 sampai dengan gambar 6 merupakan contoh sebagian hasil dari pencocokkan klasifikasi Vader default dengan kamus lexicon yang menunjukkan adanya perbedaan terkait dengan pelabelan kalimat, tentunya hal ini terjadi dikarenakan berbedanya penilaian atau pelabelan skor kata pada library Vader sentimen dengan kamus lexicon.

| | |
|---|----------|
| ['haru', 'understanding', 'free', 'papua', 'means', 'that', 'papua', 'feels', 'the', 'independence', 'of', 'an', 'independent', ... | Positive |
| ['access', 'to', 'export', 'handling', 'original', 'papua', 'papua', 'merdeka', 'to', 'august', 'under', 'the', 'auspice... | Neutral |
| ['typical', 'supporting', 'separatist', 'polyter', 'any', 'issue', 'seeking', 'attention', 'to', 'heat', 'the', 'issue', 'of', 'papua', 't... | Neutral |
| ['interpreted', 'by', 'papua', 'merdeka', 'merdeka', 'welfare', 'prosperity', 'of', 'the', 'rich', 'people', 'of', 'papua', 'develo... | Positive |
| ['lyodra', 'singing', 'papua', 'regional', 'songs', 'pekik', 'merdeka', 'belas', 'belaking', 'sembang', 'independence', 'of', ... | Negative |
| ['access', 'to', 'export', 'handling', 'original', 'papua', 'papua', 'merdeka', 'to', 'august', 'under', 'the', 'auspice... | Neutral |

GAMBAR 4

(HASIL KLASIFIKASI LEXICON PADA KELAS SENTIMEN POSITIF VADER DEFAULT)

| | |
|--|----------|
| [believe', 'christian', 'organizations', 'supports', 'the', 'movement', 'of', 'free', 'papua', 'love', 'papua', 'suffering', ... | Positive |
| [teacher', 'of', 'free', 'papua', 'organization', 'religious', 'humanitarian', 'criminals', 'teach', 'killing', 'papua', 'teac... | Neutral |
| [smart', 'papuan', 'children', 'are', 'afraid', 'of', 'being', 'fooled', 'by', 'the', 'take', 'guise', 'of', 'the', 'take', 'papua', ... | Negative |
| [wood', 'was', 'forced', 'to', 'kill', 'destroy', 'the', 'village', 'alias', 'a', 'group', 'of', 'terrorist', 'papua', 'merdeka', 'or... | Neutral |
| [papua', 'understanding', 'free', 'papua', 'the', 'struggle', 'of', 'terrorists', 'the', 'struggle', 'for', 'the', 'power', 'of', '... | Negative |

GAMBAR 5

(HASIL KLASIFIKASI LEXICON PADA KELAS SENTIMEN NEGATIF VADER DEFAULT)

| | |
|---|----------|
| [brought', 'by', 'the', 'indonesian', 'sabang', 'merauke', 'merdeka', 'negara', 'fights', 'for', 'the', 'true', 'independence', 'of', 'the', 'p... | Neutral |
| [the', 'free', 'papua', 'organization', 'killing', 'performs', 'a', 'criminal', 'action', 'of', 'the', 'original', 'facial', 'of', 'papua', 'free', 'the', '... | Negative |
| [publishing', 'the', 'issue', 'of', 'papuan', 'refugee', 'deliberately', 'when', 'the', 'republic', 'of', 'indonesia', 'was', 'celebrating', 'the', '... | Negative |
| [provoked', 'by', 'the', 'west', 'papua', 'national', 'committee', 'succeeded', 'in', 'rioting', 'wamena', 'resulting', 'in', 'a', 'spokesman... | Neutral |
| [see', 'how', 'sickening', 'the', 'separatist', 'do', 'the', 'social', 'media', 'propaganda', 'panic', 'by', 'the', 'representatives', 'of', 'the', '... | Negative |
| [hope', 'that', 'friends', 'are', 'provoked', 'by', 'the', 'west', 'papua', 'national', 'committee', 'successfully', 'rioting', 'wamena', 'due', '... | Positive |
| [see', 'how', 'sickening', 'the', 'separatist', 'do', 'the', 'social', 'media', 'propaganda', 'panic', 'by', 'the', 'representatives', 'of', 'the', '... | Negative |

GAMBAR 6

(HASIL KLASIFIKASI LEXICON PADA KELAS SENTIMEN NETRAL VADER DEFAULT)

Kemudian pada gambar 7 sampai dengan gambar 9 merupakan jumlah hasil klasifikasi leksikon dan *Vader* dengan batas skor *default* -0.5 dan 0.5 yang telah dicocokkan dengan kamus *lexicon*. Hasil dari semua klasifikasi pada gambar dibawah ini akan digunakan ke dalam rumus perhitungan pada tahap pengujian atau evaluasi model yang nantinya akan menghasilkan nilai akurasi untuk mengetahui seberapa bagus dan efektif *VADER* sentimen digunakan pada penelitian ini.

| | |
|-----------------|------------|
| Positive | 713 |
| Neutral | 50 |
| Negative | 17 |

GAMBAR 7

(HASIL KLASIFIKASI LEXICON KELAS POSITIF)

| | |
|-----------------|------------|
| Negative | 169 |
| Neutral | 19 |
| Positive | 10 |

GAMBAR 8

(HASIL KLASIFIKASI LEXICON KELAS NEGATIF)

| | |
|-----------------|------------|
| Neutral | 518 |
| Positive | 164 |
| Negative | 161 |

GAMBAR 9

(HASIL KLASIFIKASI LEXICON KELAS NETRAL)

B. Evaluasi Data

Evaluasi data pada penelitian ini dibantu dengan menggunakan tabel 3 x 3 *confusion matrix*. *Confusion matrix* merupakan metode evaluasi visual yang biasanya digunakan dalam proses klasifikasi data. Kolom *Confusion Matrix* berisikan hasil kelas prediksi, dan baris berisikan hasil kelas yang sebenarnya. Tabel matriks terdiri dari data aktual dan prediksi. Data aktual yang digunakan pada penelitian ini yaitu kamus *lexicon*, sedangkan untuk data prediksinya yaitu *Vader* sentiment. Pada Tabel 3 *confusion matrix*, rumus AA, BB, dan CC merupakan *true label* dari masing-masing kelas sentimen. AA yang berarti true positif, BB yang berarti true netral, dan CC yang berarti true negatif. *True label* menandakan bahwa hasil dari klasifikasi *VADER* dengan kamus leksikon manual

memiliki hasil yang sama. Pada rumus perhitungan berdasarkan referensi dari penelitian syukur dan istiqomah [7] [26], *true label* akan dijumlahkan yang kemudian akan dilakukan pembagian dengan keseluruhan jumlah data tweet. Berikut adalah tabel perhitungan klasifikasi sentimen *VADER* dengan kamus leksikon manual dan juga implementasi perhitungannya.

TABEL 3
(CONFUSION MATRIX)

| | | Vader Sentiment | | |
|---------------|---------|-----------------|--------|---------|
| | | Positif | Netral | Negatif |
| Lexicon Based | Kelas A | AA | AB | AC |
| | Kelas B | BA | BB | BC |
| | Kelas C | CA | CB | CC |

Keterangan:

Akurasi

$$= \frac{AA + BB + CC}{AA + AB + AC + BA + BB + BC + CA + CB + CC}$$

TABEL 4

(IMPLEMENTASI PERHITUNGAN CONFUSION MATRIX VADER DEFAULT)

| | | Vader Sentiment | | |
|---------------|---------|-----------------|--------|---------|
| | | Positif | Netral | Negatif |
| Lexicon Based | Positif | 713 | 164 | 10 |
| | Netral | 50 | 518 | 19 |
| | Negatif | 17 | 161 | 169 |

Keterangan:

$$Akurasi = \frac{713+518+169}{713+50+17+164+518+161+10+19+169} = \frac{1400}{1821} = 0,768 = 76,8$$

- positif *vader* = 780 positif *lexicon* = 887
akurasi ketepatan sentimen positif = 91,4%
- netral *vader* = 843 netral *lexicon* = 587
akurasi ketepatan sentimen netral = 61,4 %
- negatif *vader* = 198 negatif *lexicon* = 347
akurasi ketepatan sentimen negatif = 85,3 %

Di kelas sentimen negatif, *VADER* hanya menghasilkan 198 teks, sedangkan pelabelan manual berdasarkan kamus *lexicon* mengklasifikasikan lebih banyak dengan 347 teks sentimen negatif. Berbeda dengan kelas netral, *VADER* menghasilkan klasifikasi netral sebanyak 843, lebih banyak dibandingkan pelabelan manual yang mengklasifikasikan 587 teks sebagai sentimen netral. Berdasarkan perhitungan pada Tabel 4, diketahui hasil klasifikasi sentimen menggunakan *VADER* sentimen dengan polaritas skor batas *default* (batas negatifnya -0.5 dan positifnya 0.5) memberikan akurasi secara keseluruhan sebesar 76,8 %. Sedangkan pada klasifikasi *Vader* menggunakan batas nilai polaritas yang di *custom* (batas positifnya 0.4 dan negatifnya -0.3) menghasilkan akurasi yang lebih tinggi yaitu sebesar 80.7 %, dan pada tabel 5 jika dihitung masing-masing akurasi di setiap kelas sentimennya pun hasilnya lebih merata dibandingkan dengan hasil klasifikasi *VADER*

menggunakan batas skor polaritas *default* yang ternyata hasil akurasi dari kelas sentimen netralnya sangat kecil. Ini menunjukkan semakin jauh jarak antara batas polaritas skor positif dengan skor negatif, maka akan berpotensi menghasilkan akurasi yang rendah. Pada intinya nilai akurasi menjadi acuan untuk mengukur seberapa akurat sistem untuk dapat mengklasifikasikan data dengan benar.

TABEL 5
(IMPLEMENTASI PERHITUNGAN *CONFUSION MATRIX VADER CUSTOM*)

| Vader Sentiment | | | | |
|-----------------|---------|---------|--------|---------|
| Lexicon Based | | Positif | Netral | Negatif |
| | Positif | 764 | 97 | 26 |
| | Netral | 71 | 476 | 40 |
| | Negatif | 24 | 92 | 231 |

Keterangan:

$$\text{Akurasi} = \frac{764+476+231}{764 + 71 + 24 + 97 + 476 + 92 + 26 + 40 + 231} = \frac{1471}{1821} = 0,807 = 80,7\%$$

1. positif *vader* = 859 positif *lexicon* = 887
akurasi sentimen positif = 88,9%
2. netral *vader* = 665 netral *lexicon* = 587
akurasi sentimen netral = 71,5 %
3. negatif *vader* = 297 negatif *lexicon* = 347
akurasi sentimen negatif = 77,7 %

V. KESIMPULAN

Dari hasil klasifikasi yang telah dilakukan, maka dapat disimpulkan bahwa hasil analisa sentimen yang diperoleh dari 1.821 data tweet berdasarkan proses *memparsing data* dan juga menggunakan metode pendekatan *lexicon* berbasis *VADER* dengan batas skor polaritas default adalah 780 positif, 843 netral, dan 198 negatif. Jadi dapat disimpulkan opini masyarakat terhadap konflik Kelompok Kriminal Bersenjata di Papua memiliki sentimen positif 42,8 %, 46,2 % netral, dan 11 % negatif. Dan hasil analisa sentimen yang dilakukan dengan melakukan perhitungan pada proses evaluasi model menggunakan tabel *confusion matrix* mendapatkan nilai akurasi tertinggi yaitu sebesar 80,7 % pada *VADER custom* dengan batas skor polaritas negatif -0,3 dan batas positif 0,4. Hasil dari akurasi tersebut sudah termasuk tinggi, artinya pendekatan leksikon berbasis *VADER* ini cukup akurat digunakan pada penelitian ini.

REFERENSI

- [1] Z. F. Nurhadi, "Model Komunikasi Sosial Remaja Melalui Media Twitter," J. ASPIKOM, vol. 3, no. 3, p. 539, 2017, doi: 10.24329/aspikom.v3i3.154.
- [2] L. J. H. L. Mutiara Zaskya, Antonius Boham, "Twitter Sebagai Media Mengungkapkan Diri Pada Kalangan Milenial Twitter as a Media for Self-Disclosure Among Millennials," pp. 1–8, 2020.
- [3] Kominfo.go.id. (2013, 7 November). Pengguna Internet di Indonesia 63 Juta Orang. Diakses pada 6 Juli 2022, dari <https://kominfo.go.id/index.php/content/detail/3415/Kominfo+%3A+Pengguna+Internet+di+>

[Indonesia+63+Juta+orang/0/berita_satker, November 2013.](https://indonesia+63+Juta+orang/0/berita_satker_November_2013)

- [4] F. F. Mailo and L. Lazuardi, "Analisis Sentimen Data Twitter Menggunakan Metode Text Mining Tentang Masalah Obesitas di Indonesia," J. Inf. Syst. Public Heal., vol. 4, no. 1, pp. 28–36, 2019.
- [5] R. P. Hadi, "Tantangan Keamanan Nasional Indonesia dalam Menyikapi Transformasi Separatisme dan Cyberprotest di Papua," Jt. Semin. Public Int. Aff. UUM-Unair, vol. 4, no. September, pp. 1–10, 2016, [Online]. Available: http://repository.unair.ac.id/68175/3/jurnal_rany.pdf
- [7] A. Syakur, "Implementasi Metode Lexicon Base Untuk Analisis Sentimen Kebijakan Pemerintah Dalam Pencegahan Penyebaran Virus Corona Covid-19 Pada Twitter," J. Ilm. Inform. Komput., vol. 26, no. 3, pp. 247–260, 2021, doi: 10.35760/ik.2021.v26i3.4720.
- [14] M. Php, K. Bekasi, R. Noviana, J. Margonda, R. No, and P. Cina, "Sosial Twitter Dengan Metode Valence Aware Dictionary and Sentiment Reasoner (VADER) Pendahuluan," vol. 20, pp. 1–14, 2021.
- [15] E. A. Marwa and A. B. Kristanto, "Analisis Sentimen Pengungkapan Informasi Manajemen : Text Mining Berbasis Metode VADER," vol. 6, pp. 2973–2984, 2022.
- [20] W. Medhat, A. Hassan and H. Korashy, "Sentiment analysis algorithms and applications," Ain Shams Engineering Journal, vol. 1, pp. 1093-1113, 2013.
- [23] Minqing Hu and Bing Liu. "Mining and Summarizing Customer Reviews." Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD-2004), Aug 22-25, 2004, Seattle, Washington, USA.
- [24] A.- Arini, L. K. Wardhani, and D.- Octaviano, "Perbandingan Seleksi Fitur Term Frequency & Tri-Gram Character Menggunakan Algoritma Naïve Bayes Classifier (Nbc) Pada Tweet Hashtag #2019gantipresiden," Kilat, vol. 9, no. 1, pp. 103–114, 2020, doi: 10.33322/kilat.v9i1.878.
- [26] L. Istiqomah, "Analisis sentimen penerimaan CPNS menggunakan pendekatan vader," *Digilib.Uin-Suka.Ac.Id*, no. April 2020, p. 506812, 2021, [Online]. Available: http://digilib.uin-suka.ac.id/1739711/BAB_I,_V,_DAFTAR_PUSTAKA.pdf
- [28] James Suciadi, "Studi Analisis Metode-Metode Parsing Dan Interpretasi Semantik Pada Natural Language Processing," J. Inform., vol. 2, no. 1, pp. 13–22, 2001, [Online]. Available: http://puslit2.petra.ac.id/_journal/index.php/inf/article/view/15799