

ANALISIS SENTIMEN BAHASA INDONESIA PADA TWITTER MENGGUNAKAN STRUKTUR *TREE* BERBASIS LEKSIKON

Feby Tri Saputra*¹, Yani Nurhadryani², Sony Hartono Wijaya³, Defina⁴

^{1,2,3} Departemen Ilmu Komputer, Fakultas Matematika dan Ilmu Pengetahuan Alam, Institut Pertanian Bogor

⁴ Departemen Ilmu Keluarga dan Konsumen Fakultas Ekologi Manusia, Institut Pertanian Bogor

Email: ¹saputra_feby@apps.ipb.ac.id, ²yani_nurhadryani@apps.ipb.ac.id, ³sony@apps.ipb.ac.id,

⁴defina@apps.ipb.ac.id

(Naskah masuk: 21 September 2020, diterima untuk diterbitkan: 02 Februari 2021)

Abstrak

Jumlah opini di media sosial seperti Twitter tersebar luas sehingga tidak mungkin membaca semua opini untuk mendapatkan seluruh sentimen. Analisis sentimen merupakan salah satu metode untuk mengatasi masalah tersebut. Salah satu pendekatan dalam analisis sentimen adalah berbasis leksikon. Pendekatan berbasis leksikon dapat menghasilkan performa yang baik pada lintas topik pembicaraan tanpa memerlukan pelatihan data. Namun, pendekatan berbasis leksikon sangat bergantung pada kelengkapan dan keragaman sentimen leksikon. Selain itu, hubungan antarkata sangat penting untuk diperhatikan karena dapat mengubah polaritas sentimen pada teks. Hubungan antarkata dapat direpresentasikan dengan baik menggunakan struktur *tree*. Penelitian ini menggunakan struktur *tree* sebagai interpretasi hubungan antarkata dalam pembentukan kalimat dengan menambahkan kata ke dalam sentimen leksikon. Metode berbasis *tree* diujikan pada data dengan lintas topik seperti data *twit* Pilgub Jabar 2018, Pilpres 2019, dan pandemik COVID-19. Ketiga data uji memiliki proporsi kelas yang tidak seimbang, dengan kelas terbanyak merupakan kelas positif. Metode berbasis *tree* menghasilkan akurasi sebesar 64,97% (meningkat 1,26%) pada data Pilgub Jabar 2018, 64,33% (meningkat 11,41%) pada data Pilpres 2019, dan 66,24% (meningkat 7,61%) pada data pandemik COVID-19. Metode berbasis *tree* dapat menghasilkan akurasi yang stabil pada beberapa lintas topik dibuktikan dengan standar deviasi akurasi yang kecil (0,97%) bahkan lebih kecil dari metode tanpa *tree* (5,4%). Metode berbasis *tree* dapat meningkatkan *weighted f1-measure* pada data Pilpres 2019 sebesar 10,45% dan data pandemik COVID-19 sebesar 8,1%, sedangkan hasil pada data Pilgub 2018 tidak berbeda secara signifikan. Hasil akurasi dan *weighted f1-measure* memiliki selisih yang kecil sehingga pengukuran akurasi valid dan tidak bias terhadap data tidak seimbang.

Kata kunci: *analisis sentimen, hubungan antarkata, lexicon-based, struktur tree, twitter*

INDONESIA SENTIMENT ANALYSIS ON TWITTER USING A LEXICON-BASED TREE CONSTRUCTION

Abstract

The number of opinions on social media like Twitter is so widespread that it's impossible to read all those opinions to get all the sentiments. Sentiment analysis is one of the methods that could overcome this problem. The lexicon-based approach is one of the sentiment analysis approaches which perform well across data topics without training. However, the lexicon-based approach relies heavily on the completeness and diversity of sentiment lexicons. The relationship between words is important because it could change the sentiment polarity in the text. The tree structure could represent the relationship between words well. This study uses a tree structure as an interpretation of the relationship between words in a sentence. The tree structure is constructed by adding words to the lexicon sentiment. The tree-based method is tested on cross-topic data such as the tweet data of the 2018 West Java Governor Election, the 2019 Presidential Election, and the COVID-19 pandemic. All data used has an unbalanced class proportion, with the positive class being dominant. The accuracy results of the tree-based method on all data consecutively are 64.97% (increased by 1.26%), 64.33% (increased by 11.41%), and 66.24% (increased by 7.61%). The tree-based method produce stable accuracy on several topics proved by the small accuracies standard deviation (0.97%) that even smaller than the non-tree method (5.4%). The weighted *f1-measure* increases of the tree-based method on all data consecutively are 0% (equal), 10.45%, and 8.1%. The small difference between the weighted *f1-measure* and accuracy concludes that the accuracy resulted is valid.

Keywords: *lexicon-based, sentiment analysis, tree structure, twitter, words relationship*

1. PENDAHULUAN

Jumlah opini publik yang tersebar di berbagai media semakin meningkat seiring berkembangnya teknologi, informasi, dan komunikasi. Melihat hal tersebut, tidak mungkin untuk membaca setiap opini yang disampaikan satu per satu untuk mendapatkan sentimen yang disampaikan oleh masyarakat. Analisis sentimen atau biasa disebut *opinion mining* dapat dijadikan suatu solusi untuk mendapatkan sentimen publik.

Pada era digital seperti sekarang ini, media sosial telah menjadi salah satu wadah yang sering digunakan masyarakat dalam mengutarakan opininya. Data pada tahun 2018 menunjukkan bahwa pengguna internet Indonesia telah mencapai 64,79% dari total penduduk Indonesia yang berjumlah 264,16 juta jiwa. Adapun jumlah pengguna internet Indonesia adalah 171,17 juta jiwa. Adapun layanan media sosial termasuk ke dalam layanan nomor dua yang sering diakses dengan persentase pengguna sebesar 18,9% (APJII, 2018). Salah satu media sosial yang berpotensi menyebarkan informasi secara cepat dan meluas adalah Twitter (Kwak dkk., 2010). Hal ini didukung oleh hasil survei We are social pada bulan Januari 2020 yang menunjukkan bahwa Twitter menduduki peringkat ke-5 pengguna terbanyak di Indonesia setelah YouTube, WhatsApp, Facebook, dan Instagram (Simon, 2020).

Analisis sentimen dapat mengklasifikasikan teks yang mengandung opini sebagai kelas positif, negatif, atau netral secara otomatis (Indurkha dan Damerau, 2010). Secara umum, pendekatan dalam analisis sentimen terbagi ke dalam dua hal, yaitu pendekatan berbasis *corpus* dan pendekatan berbasis leksikon. Pendekatan berbasis *corpus* memiliki akurasi yang lebih tinggi dibandingkan dengan pendekatan berbasis leksikon (Abdulla dkk., 2013). Namun, pendekatan berbasis *corpus* sangat bergantung pada kualitas dan jumlah data pelatihan (Pamungkas dan Putri, 2016). Akurasi pada pendekatan berbasis leksikon dapat ditingkatkan dengan cara memperluas leksikon dengan mempertimbangkan dialek dan spesifik tema pembicaraan, menambahkan rincian bobot pada sentimen positif dan negatif, meningkatkan kombinasi atau agregasi dari polaritas setiap kata untuk menghitung polaritas keseluruhan teks, melakukan eksperimen yang lebih menyeluruh pada *corpus* yang lebih besar dan beragam, serta penanganan sarkasme pada suatu teks (Abdulla dkk., 2014). Selain itu, metode berbasis leksikon menghasilkan performa yang baik pada lintas topik pembicaraan, dapat dengan mudah ditingkatkan dengan berbagai sumber pengetahuan, dan tidak memerlukan pelatihan lebih lanjut (Taboada dkk., 2011; Palanisamy, Yadav, dan Elchuri, 2013).

Analisis sentimen berbasis leksikon pada bahasa Indonesia telah dilakukan oleh Pamungkas dan Putri (2016) dengan melakukan analisis sentimen komentar pengguna pada Google PlayStore dan Apple AppStore. Penelitian tersebut menggunakan sentimen leksikon

Inggris (SentiWordNet) sehingga perlu dilakukan penerjemahan teks ke dalam bahasa Inggris terlebih dahulu. SentiWordNet memiliki lebih dari 100 000 kata berikut makna polisemi dan skor polaritasnya (positif dan negatif) sehingga banyak digunakan di berbagai studi analisis sentimen (Pamungkas dan Putri, 2016; Baccianella, Esuli, dan Sebastiani, 2010). Akurasi yang dihasilkan pada penelitian tersebut sebesar 68%. Kekurangan pada penelitian Pamungkas dan Putri (2016) adalah metode yang digunakan tidak dapat menangani bahasa tidak baku, fenomena *thwarted expectation*, dan kata yang bersifat polisemi sehingga sulit menentukan makna yang tepat dalam kalimat. Adapun fenomena *thwarted expectation* terjadi ketika suatu teks menjelaskan ekspektasi yang selanjutnya menjelaskan kekecewaan terhadap suatu hal.

Penelitian berikutnya tentang analisis sentimen berbasis leksikon pada bahasa Indonesia dilakukan oleh Lailiyah, Sumpeno, dan Purnama (2017) dengan melakukan analisis sentimen pada data pengaduan publik di Twitter dan situs pemerintah. Lailiyah, Sumpeno, dan Purnama (2017) membandingkan akurasi pada penggunaan dua *lexical resource*, yaitu sentimen leksikon Indonesia dan SentiWordNet. Akurasi yang dihasilkan sentimen leksikon Indonesia lebih tinggi dibandingkan dengan SentiWordNet ketika diujikan pada data berbahasa Indonesia. Akurasi yang dihasilkan meningkat dari 47% menjadi 65.4% pada data Twitter dan meningkat dari 56.85% menjadi 81.4% pada data situs pemerintah. Namun, penelitian Lailiyah, Sumpeno, dan Purnama (2017) masih belum mempertimbangkan pengingkaran dalam suatu kalimat.

Penelitian Saputra dan Nurhadryani (2018) telah meningkatkan akurasi dari penelitian Lailiyah, Sumpeno, dan Purnama (2017) dengan mempertimbangkan pengingkaran dan struktur gramatikal kalimat dalam bahasa Indonesia. Penanganan struktur gramatikal kalimat dapat mengatasi kesalahan klasifikasi pada kalimat majemuk. Akurasi yang dihasilkan meningkat dari 60.15% menjadi 63.7% ketika diterapkan pada data *twit* Pilgub Jabar 2018. Salah satu kekurangan pada penelitian Saputra dan Nurhadryani (2018) adalah sentimen leksikon Indonesia masih belum bisa membedakan kata berdasarkan kelas kata sehingga terjadi ambiguitas. Sebagai contoh, kata "bisa" dalam kata kerja artinya "dapat" yang bermakna positif, sedangkan kata "bisa" dalam kata benda merupakan zat beracun yang bermakna negatif. Namun menurut Li, Zhu, dan Zhang (2011), membedakan makna kata berdasarkan kelas kata tidak meningkatkan akurasi. Oleh karena itu, penelitian ini belum berfokus pada pembuatan sentimen leksikon Indonesia yang dapat membedakan setiap kata berdasarkan kelas kata. Kekurangan lainnya yaitu, sentimen leksikon Indonesia yang digunakan masih relatif sedikit dan perhitungan bobot hanya menghitung frekuensi kemunculan kata berdasarkan sentimen leksikon saja tanpa memperhatikan makna dalam hubungan antarkata. Adapun fitur yang digunakan

biasanya adalah bag-of-word. Hal tersebut menyebabkan hilangnya informasi pada hubungan setiap kata (Pak dan Paroubek, 2011; Li dkk., 2010). Adapun hubungan antarkata dapat mengubah polaritas sentimen pada teks. Hubungan antarkata dalam suatu teks dapat direpresentasikan dengan baik menggunakan struktur *tree* (Pak dan Paroubek, 2011). Adapun struktur *tree* dapat merepresentasikan hirarki dalam pembentukan kalimat.

Penelitian analisis sentimen berbasis *tree* sebelumnya hanya memanfaatkan struktur *tree* sebagai ekstraksi fitur yang hasilnya kemudian digunakan pada pelatihan berbasis *machine learning* dan *deep learning*. Nakagawa, Inui, dan Kurohashi (2010) menggunakan struktur *tree* sebagai *hidden variable* pada metode klasifikasi conditional random fields (CRFs) dengan rata-rata akurasi 84% pada data berbahasa Jepang dan 81.3% pada data berbahasa Inggris. Pak dan Paroubek (2011) menggunakan struktur *tree* untuk menghasilkan *feature vector* pada metode klasifikasi SVM dengan akurasi 85.1%. Penelitian menggunakan *deep learning* dilakukan oleh Teng dan Zhang (2017) menggunakan *bidirectional tree LSTM* dengan akurasi 94.8%. Ketiga penelitian tersebut menunjukkan bahwa analisis sentimen menggunakan struktur *tree* dapat menghasilkan akurasi yang cukup baik (lebih dari 80%). Akan tetapi penelitian tersebut memiliki biaya komputasi yang lebih besar dibanding analisis sentimen berbasis leksikon karena melakukan konstruksi *tree* sekaligus pelatihan data. Maka dari itu analisis sentimen berbasis leksikon menggunakan struktur *tree* perlu dilakukan untuk memperkecil biaya komputasi.

Berdasarkan permasalahan pada penelitian sebelumnya, penelitian ini bertujuan untuk meningkatkan performa analisis sentimen berbasis leksikon sebelumnya dengan menangani kata tidak baku dan menggunakan struktur *tree* sebagai interpretasi hubungan antarkata dalam suatu teks. Struktur *tree* dibangun dengan mempertimbangkan penambahan kata pada sentimen leksikon. Hasil penelitian ini diharapkan dapat menghasilkan model analisis sentimen berbasis leksikon yang memiliki performa tinggi dan stabil pada berbagai lintas topik tanpa pelatihan.

2. METODE PENELITIAN

2.1. Data Penelitian

Data yang digunakan pada penelitian ini adalah sebagai berikut:

1. *Dataset Sentiment Lexicons for 81 Languages* berdasarkan penelitian Chen dan Skiena (2014) sebagai sentimen leksikon Indonesia.
2. Daftar kata baku berdasarkan penelitian Aziz (2013). Daftar kata baku ini digunakan untuk merubah kata tidak baku menjadi kata baku.
3. Daftar kata konjungsi sebagai tanda pemisah antarklausa dari penelitian Alwi dkk. (2010).
4. Daftar tanda baca sebagai tanda pemisah antarklausa. Tanda baca yang digunakan yaitu, tanda titik, koma, titik koma, seru, tanya dan

ampersan (Wahyuni, Krishandini, dan Defina, 2018).

5. Daftar kata negasi digunakan untuk menegatifkan atau mengingkarkan predikat yang bersumber dari Alwi dkk. (2010).
6. Data uji yang digunakan sebagai berikut:
 - a. Data twit pada masa kampanye Pilgub Jawa Barat 2018 berdasarkan penelitian Saputra dan Nurhadryani (2018) sebanyak 394 twit.
 - b. Data twit yang diambil pada masa kampanye Pilpres 2019 pada tanggal 13 Februari – 16 April 2019 secara *real-time*.
 - c. Data twit yang diambil pada masa pandemik COVID-19 pada tanggal 16 April 2020 – 15 Mei 2020 secara *real-time*.

Tabel 1 menjelaskan kata kunci yang digunakan dalam pencarian data twit. Komposisi kelas sentimen pada data sampel ditunjukkan pada Tabel 2. Tabel 2 menunjukkan bahwa ketiga data sampel memiliki proporsi antar kelas yang tidak seimbang di mana kelas positif memiliki proporsi terbesar.

Tabel 1 Kata kunci pencarian data twit Pilpres 2019 dan Pandemi COVID-19 berdasarkan nama akun dan tagar

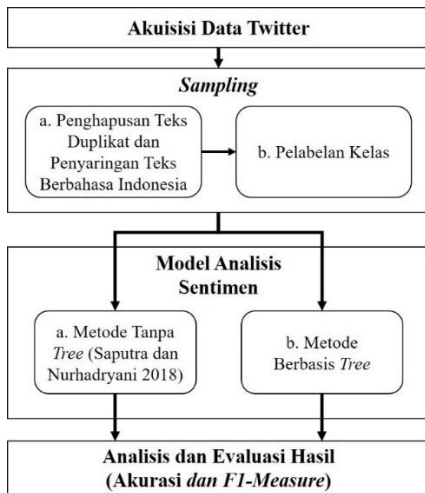
Data Twit	Kata Kunci Pencarian		
	Akun	Tagar	
Pilpres 2019			
A	Joko Widodo dan Ma'ruf Amin	@jokowi, @KHMaru fAmin_ @Jokowi_ Amin, @JokowiA min_ID	#MenujuIndonesiaMaju, #01JokowiLagi, #01IndonesiaMaju, #jokowi2periode, #indonesiamaju, #2019PilihJokowi, #JokowiHarapanPasti, #JokowiOrangBaik, #jokowimaruf, #JokowiLagi, #2019JokowiPresiden, #jokowidicintaiakyat, #kitauntuksatu, #JokowiCintaUlama
B	Prabowo Subianto dan Sandiaga Salahudd in Uno	@prabowo, @sandiuno, @Prabowo _Uno, @prabowos andi_ID	#AdilMakmur, #PrabowoSandiBerdikari, #2019AdilMakmur, #2019PrabowoSandi, #2019PrabowoPresiden, #2019GantiPresiden, #2019PrabowoPresidenRI, #PrabowoIndonesiaMenang, #2019PrabowoSandiMenang, #PrabowoSandiunoIndonesia, #PrabowoSandiSolusiNKRI, #PrabowoSandiBawaSolusi, #MilenialPilihPrabowo, #RakyatInginkanPerubahan
Pandemik COVID-19	-	-	#dirumahaja, #indonesialockdown, #psbb

Tabel 2 Distribusi kelas sentimen pada data sampel

Data	Positif	Negatif	Netral	Total
Pilgub Jabar 2018	233 (59,14%)	62 (15,74%)	99 (25,13%)	394
Pilpres 2019	3974 (45,66%)	3224 (37,04%)	1505 (17,29%)	8703
Pandemik COVID-19	249 (63,2%)	69 (17,51%)	76 (19,29%)	394

2.2. Tahapan Penelitian

Penelitian ini terdiri atas beberapa tahapan (Gambar 1).



Gambar 1 Tahapan penelitian

Tahapan penelitian ini membentuk *tree* sebagai interpretasi hubungan antarkata dalam suatu teks untuk membandingkan akurasi dengan penelitian Saputra dan Nurhadryani (2018).

Akuisisi Data Twitter

Proses akuisisi data yang dilakukan mengacu pada penelitian Saputra dan Nurhadryani (2018). Proses akuisisi data menggunakan *library* Tweepy. Twit didapatkan dengan API jenis *search* berdasarkan kata kunci pencarian yang terdiri atas nama akun dan tagar (Tabel 1). Atribut yang diambil adalah atribut “created_at”, “user.screen_name”, “full_text”, “user.location”. Atribut “full_text” yang hanya berisi tautan, tagar (#), dan mention (@) tidak akan digunakan karena tidak mengandung opini.

Sampling

Tahap ini merupakan proses pengambilan sampel yang menggambarkan karakteristik dari populasi data twitter yang telah didapatkan pada tahap sebelumnya. Tahap ini terdiri atas praproses data (penghapusan teks duplikat dan penyaringan bahasa Indonesia pada teks) serta pelabelan kelas. Hasil dari tahap ini kemudian digunakan sebagai data uji.

- a. Penghapusan Teks Duplikat dan Penyaringan Teks Berbahasa Indonesia
 Penghapusan data duplikat dilakukan dengan menggunakan fungsi “Remove Duplicates” pada Microsoft Excel dan pencarian teks bahasa Indonesia menggunakan fungsi “cld2::detect_language()” pada bahasa pemrograman R. Hal tersebut dilakukan untuk mendapatkan data twit yang berbahasa Indonesia saja. Setelah itu, dilakukan penghapusan baris yang mengandung nilai NA pada teks yang bahasanya tidak ditemukan.
- b. Pelabelan Kelas

Pelabelan kelas dilakukan secara manual oleh pakar bahasa Indonesia untuk dijadikan sebagai kelas yang diharapkan pada setiap twit yang nantinya digunakan dalam analisis dan evaluasi hasil. Label kelas yang digunakan adalah kelas positif, negatif, dan netral. Melakukan pelabelan data teks diperlukan cukup waktu jika data yang digunakan cukup besar. Untuk mengatasi masalah tersebut dapat dilakukan *sampling* sebagai representasi keseluruhan data dengan tingkat toleransi tertentu. Persamaan (1) menunjukkan salah satu perhitungan *sampling* menggunakan rumus Slovin (Umar, 2002). Rumus Slovin digunakan karena cocok digunakan pada data yang besar dan kelasnya tidak diketahui secara pasti.

$$n = \frac{N}{1 + Ne^2} \tag{1}$$

Di mana :

n = ukuran sampel

N = ukuran populasi

e = margin of error (toleransi ketelitian)

Model Analisis Sentimen

Data uji yang dihasilkan pada tahap *sampling* kemudian dimodelkan dengan dua model analisis sentimen yang berbeda. Hal ini dilakukan untuk membandingkan performa yang dihasilkan dari kedua model tersebut. Model analisis sentimen yang dibandingkan pada penelitian ini, yaitu metode tanpa *tree* dan metode berbasis *tree*.

a. Metode Tanpa *Tree*

Tahapan ini menggunakan metode pada penelitian Saputra dan Nurhadryani (2018). Adapun fitur yang diekstrak adalah unigram yang mengandung sentimen positif atau negatif berdasarkan sentimen leksikon Indonesia dengan mempertimbangkan kata negasi, konjungsi, dan tanda baca. Perhitungan bobot pada sentimen leksikon Indonesia diperoleh dari frekuensi kemunculan kata yang sesuai dengan sentimen leksikon Indonesia, jika yang ditemukan kata positif maka skor positif bertambah satu poin, hal tersebut pun berlaku pada kata negatif.

Klasifikasi pada fitur yang diperoleh berdasarkan sentimen leksikon Indonesia hanya membandingkan frekuensi kemunculan kata positif dan kata negatif seperti pada penelitian Lailiyah, Sumpeno, dan Purnama (2017). Jika frekuensi kata positif lebih tinggi maka hasil klasifikasinya adalah kelas positif, hal ini berlaku sebaliknya. Jika frekuensi kata positif sama dengan kata negatif maka hasil klasifikasinya adalah kelas netral.

b. Metode Berbasis *Tree*

Tahapan ini diusulkan oleh peneliti dengan menggunakan struktur *tree* sebagai interpretasi hubungan antarkata dalam suatu teks. Metode berbasis *tree* terdiri atas beberapa tahapan (Gambar 2).



Gambar 2 Tahapan metode berbasis tree

1. Praproses Data

Praproses data yang dilakukan terdiri atas pembersihan data, tokenisasi data, mengubah kata menjadi kata baku menggunakan data penelitian Aziz (2013), dan POS Tagging (Gambar 3).



Gambar 3 Tahapan praproses data

Pembersihan data yang dilakukan terdiri atas:

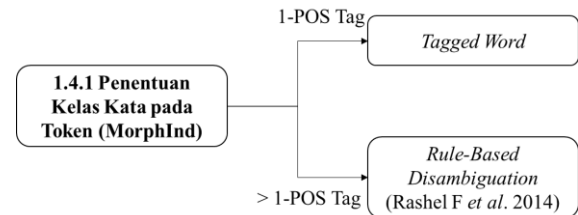
- Penghapusan karakter yang tidak diperlukan seperti simbol, angka, *mention*, tautan, *emoticon*, dan karakter “RT” karena hal tersebut tidak mengandung sentimen. Namun, simbol yang terdapat pada daftar tanda baca yang bersumber dari Wahyuni, Krishandini, dan Defina (2018) tidak dihapuskan karena tanda baca tersebut digunakan sebagai pemisah antarklausa. Selain itu, tagar (#) tidak dihapuskan karena terdapat kemungkinan tagar memiliki sentimen seperti contoh #Asyik, #Merdeka, dll.
- Penghapusan kalimat tanya karena kalimat tersebut dianggap netral karena secara eksplisit, suatu pertanyaan tidak memberikan penilaian terhadap suatu hal (dianggap netral).
- Perubahan format teks menjadi huruf kecil.

```

{
  '0': {'sentimen_score': 0,
        'kelas_kata': ['NN'],
        'value': 'pihak'},
  '1': {'sentimen_score': 0,
        'kelas_kata': ['PRP'],
        'value': 'nya'},
  '2': {'sentimen_score': 1,
        'kelas_kata': ['VB'],
        'value': 'berkomitmen'},
  '3': {'sentimen_score': 0,
        'kelas_kata': ['SC'],
        'value': 'untuk'},
  '4': {'sentimen_score': -1,
        'kelas_kata': ['VB'],
        'value': 'menghapus'},
  '5': {'sentimen_score': 0,
        'kelas_kata': ['NN'],
        'value': 'komentar'},
  '6': {'sentimen_score': -1,
        'kelas_kata': ['NEG'],
        'value': 'tidak'},
  '7': {'sentimen_score': 1,
        'kelas_kata': ['JJ'],
        'value': 'pantas'},
}
    
```

Gambar 4 Hasil praproses data

Tahapan selanjutnya adalah menghasilkan token dalam bentuk *dictionary* dengan atribut "no_id", "sentimen_score" sebagai *positive negative polarity* (PN polarity), "value" sebagai kata, dan "kelas_kata" sebagai kelas kata yang setiap katanya sudah diubah menjadi kata baku (Gambar 4). Kelas kata didapatkan menggunakan metode POS tagging pada penelitian Rashel dkk. (2014) yang secara umum dapat ditunjukkan pada Gambar 5. Adapun kata pada token telah diubah menjadi kata baku.



Gambar 5 Tahapan POS Tagging pada penelitian Rashel dkk. (2014)

2. Penambahan Data Leksikon dan Data Pembakuan Kata

Setiap kata yang telah dipecah pada praproses data dilakukan pelabelan manual untuk menentukan polaritas setiap kata yang diverifikasi oleh pakar bahasa Indonesia. Penentuan polaritas ini dijadikan sebagai acuan untuk menambahkan kata-kata pada sentimen leksikon Indonesia. Skema pada penambahan data leksikon dibagi menjadi dua, yaitu hanya menambahkan leksikon dan menambahkan sekaligus memvalidasi leksikon. Validasi yang dilakukan adalah dengan mengubah polaritas kata pada sentimen leksikon ke polaritas yang sebenarnya, sebagai contoh kata pada sentimen leksikon negatif yang dilabeli sebagai kata positif maka dilakukan penghapusan kata tersebut pada sentimen leksikon negatif dan ditambahkan pada sentimen leksikon positif. Jika kata pada sentimen leksikon negatif dilabeli sebagai kata netral maka hanya dilakukan penghapusan kata. Hal ini berlaku juga pada sentimen leksikon positif. Selain itu, setiap kata yang mengandung kata tidak baku dan bahasa gaul dipadankan dengan kata bakunya secara manual. Tabel 3 menunjukkan contoh validasi dan pepadanan kata tidak baku ke kata baku.

Tabel 3 Contoh validasi dan pepadanan kata tidak baku ke kata baku

kata	Kelas kata	Leksikon	Validasi opini	Kata baku	Evaluasi
terselesaikan	X	Negatif	Positif		benar
memainkan	VB	Negatif	Netral		benar
tergantikan	VB	Positif	Negatif		benar
tanda	NN	Positif	Netral		benar
manyep	X	Netral	Positif	mantap	benar

3. Ekstraksi Fitur

Setiap token yang dihasilkan pada tahapan praproses data dijadikan ke dalam bentuk *tree*. Adapun *root node* dijadikan sebagai hasil klasifikasi, *child node* dijadikan sebagai kelas sementara, dan *leaf* sebagai *term* yang memiliki nilai *positive negative polarity* (PN *polarity*). Nilai polaritas tersebut didapatkan berdasarkan sentimen leksikon Indonesia. Algoritma yang dilakukan adalah sebagai berikut:

- Jika kalimat yang akan diklasifikasikan adalah kalimat majemuk, maka akan dipisahkan ke dalam beberapa kalimat berdasarkan tanda baca intonasi selesai, seperti “.”, “!”. Jumlah kalimat yang terpisah menunjukkan jumlah *tree* yang akan dibentuk.
- Setiap kalimat dipisahkan ke dalam beberapa klausa jika terdapat kata konjungsi maupun tanda baca intonasi jeda seperti “;”, “:”, “,” yang menghubungkannya (Gambar 6).
- Pembentukan struktur *tree* diawali dengan membuat node *root* pada setiap klausa untuk dijadikan sebagai *subtree*.
- Pembentukan struktur *tree* dilanjutkan pada setiap *subtree* dengan mengunjungi setiap *term* pada token yang telah terpisah oleh konjungsi.
- Konjungsi yang ditemukan dijadikan sebagai *node* cabang untuk menggabungkan kembali antarklausa menjadi suatu kalimat.

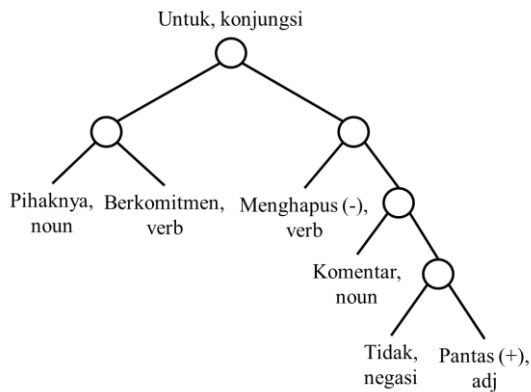
Tree yang dihasilkan oleh algoritma tersebut dapat dilihat pada Gambar 7.

Pihaknya berkomitmen (untuk, konjungsi) menghapus komentar (tidak, negasi) pantas



- Pihaknya berkomitmen
- menghapus komentar (tidak, negasi) pantas

Gambar 6 Pemisahan klausa pada token



Gambar 7 *Tree* yang dihasilkan

4. Klasifikasi Menggunakan *Rule-Based Classification*

Tahapan ini menggunakan aturan-aturan dalam pembentukan struktur *tree* sebagai interpretasi

hubungan antarkata dalam suatu teks. Setiap *leaf* pada *tree* yang dihasilkan memiliki nilai *positive negative polarity* (PN *polarity*) masing-masing. Nilai-nilai tersebut akan digabungkan untuk menentukan keseluruhan polaritas tersebut (hasil klasifikasi) dalam suatu teks berdasarkan polaritas dari hubungan antarkata. Hasil penggabungan akan disimpan pada *parent node* dari setiap *leaf* tersebut sebagai kelas sementara. Penentuan polaritas dari hubungan antarkata dilakukan dengan melakukan penelusuran *tree* dari level terendah (*bottom-up*) dengan mempertimbangkan *rule* tertentu. *Rule* yang digunakan dalam menentukan polaritas dari hubungan antarkata, yaitu:

- menggunakan sifat perkalian kedua bilangan positif dan negatif (Tabel 4).
- menggunakan sifat logika matematika seperti konjungsi untuk kata “dan” (Tabel 5) dan disjungsi untuk kata “atau” (Tabel 6).
- kata negasi memiliki nilai sentimen negatif karena kata tersebut digunakan sebagai pengingkaran dari suatu pernyataan.
- menggunakan sifat hubungan kata benda dengan kata sifat dan kata kerja (Tabel 7).
- menggunakan sifat hubungan kata kerja dengan kata sifat (Tabel 8).
- penelusuran *tree* tidak dilakukan pada *right node subtree* yang memiliki *parent* konjungsi konsesif.
- penelusuran *tree* tidak dilakukan pada *left node subtree* yang memiliki *parent* konjungsi “tetapi”.

Ilustrasi penelusuran *tree* yang dihasilkan dapat dilihat pada Gambar 8.

Tabel 4 *Rule* dalam menentukan polaritas hubungan antarkata menggunakan sifat perkalian dua bilangan

Kata 1	Kata 2	Kelas
+	+	+
+	-	-
-	+	-
-	-	+

Tabel 5 *Rule* dalam menentukan polaritas hubungan antarkata menggunakan sifat logika matematika konjungsi

Kata 1	Kata 2	Kelas
+	+	+
+	-	-
-	+	-
-	-	-

Tabel 6 *Rule* dalam menentukan polaritas hubungan antarkata menggunakan sifat logika matematika disjungsi

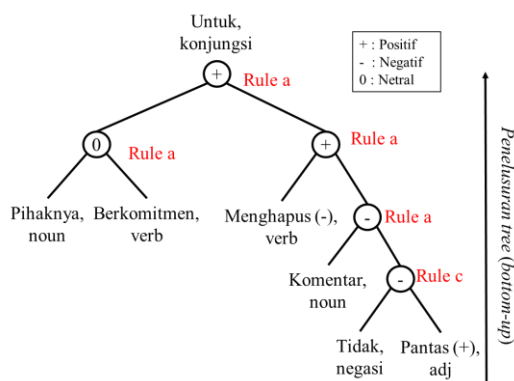
Kata 1	Kata 2	Kelas
+	+	+
+	-	+
-	+	+
-	-	-

Tabel 7 Rule dalam menentukan polaritas hubungan antarkata menggunakan sifat hubungan kata benda dengan kata sifat dan kata kerja

Kata 1 (NN)	Kata 2 (JJ or VB)	Kelas
-	+	-
-	-	-
+	+	+
+	-	-

Tabel 8 Rule dalam menentukan polaritas hubungan antarkata menggunakan sifat hubungan kata kerja dengan kata sifat

Kata 1 (VB)	Kata 2 (JJ)	Kelas
-	+	-
-	-	-
+	+	+
+	-	-



Gambar 8 Penelusuran tree dalam penentuan kelas (klasifikasi)

Analisis dan Evaluasi Hasil

Tahapan ini mengukur performa dari kedua pemodelan klasifikasi pada ketiga data uji yang digunakan. Pengukuran ini biasanya menggunakan perhitungan akurasi (Han, Kamber, dan Pei, 2012). Pengukuran performa yang digunakan pada penelitian ini adalah pengukuran akurasi dan *f1-measure*.

F1-measure digunakan untuk membandingkan tingkat akurasi ke dalam satu nilai berdasarkan nilai *precision* dan *recall*. *Precision* dapat dianggap sebagai ukuran ketepatan yang menunjukkan persentase ketepatan sistem dalam memprediksi data sesuai kelasnya, sedangkan *recall* adalah ukuran kelengkapan yang menunjukkan seberapa lengkap hasil relevan yang ditampilkan oleh sistem.

Pengukuran *precision*, *recall*, akurasi, dan *f1-measure* berdasarkan Persamaan (2), (3), (4), dan (5). Pengukuran *precision*, *recall*, dan *f1-measure* dilakukan pada setiap kelas sehingga setiap pengukuran memiliki tiga nilai. Keseluruhan skor *f1-measure* dihitung berdasarkan *weighted f1-measure* yang menjadikan setiap ukuran sampel kelas sebagai bobot. *Weighted f1-measure* dipilih karena analisis sentimen bersifat *multi-class* dan jumlah data antar kelas yang digunakan tidak seimbang. Pengukuran akurasi, *precision*, *recall*, dan *f1-measure* dihitung berdasarkan tabel *Confusion Matrix*. Persamaan (2) menunjukkan perhitungan akurasi. Terdapat empat

istilah pada tabel *Confusion Matrix*, yaitu sebagai berikut:

1. TP (*True Positives*) adalah data aktual positif diprediksi benar oleh klasifikasi.
2. FP (*False Positives*) adalah data aktual negatif diprediksi salah oleh klasifikasi.
3. FN (*False Negatives*) adalah data aktual positif diprediksi salah oleh klasifikasi.
4. TN (*True Negatives*) adalah data aktual negatif diprediksi benar oleh klasifikasi.

$$Precision = \frac{TP}{TP+FP} \tag{2}$$

$$Recall = \frac{TP}{TP+FN} \tag{3}$$

$$Akurasi = \frac{TP+TN}{TP+TN+FP+FN} \tag{4}$$

$$F1-Measure = \frac{2*(Precision*Recall)}{Precision+Recall} \tag{5}$$

3. HASIL DAN PEMBAHASAN

3.1. Akuisisi Data Twitter

Data twit tentang Pilpres 2019 telah terkumpul 187450 data. Data twit tentang pandemik COVID-19 telah terkumpul 16374 data twit.

3.2. Sampling

Jumlah kedua data hasil *sampling* dapat dilihat pada Tabel 9. Jumlah sampel pada data Pilpres 2019 menggunakan metode Slovin dengan tingkat toleransi sebesar 1%, sedangkan pada data pandemik COVID-19 hanya menggunakan *random sampling* saja dengan jumlah yang sama dengan data Pilgub Jabar 2018 karena data ini dijadikan sebagai data uji tambahan.

Tabel 9 Jumlah data hasil *sampling*

Tahap Sampling	Data Twit		
	Pilgub Jabar 2018	Pilpres 2019	Pandemik COVID-19
Data Awal	-	187450	16374
Setelah Penghapusan Teks Duplikat	-	120713	11707
Data Teks Berbahasa Indonesia	-	106753	10026
Data Tanpa NA	-	67089	7793
Jumlah Sampel	394	8703	394

3.3. Model Analisis Sentimen

Model analisis sentimen diimplementasikan menggunakan Python 3. Pada metode berbasis *tree* dilakukan penambahan data leksikon dan pembakuan kata, sedangkan pada metode tanpa *tree* tidak.

Data sampel yang digunakan untuk penambahan data leksikon dan pembakuan kata adalah data twit

tentang Pilpres 2019 yang berjumlah 8703. Jumlah kata yang dihasilkan dari data sampel tersebut 15435. Tabel 10 menjelaskan jumlah penambahan kata tiap kelas sentimen pada sentimen leksikon berdasarkan data sampel.

Tabel 10 Jumlah penambahan kata pada sentimen leksikon berdasarkan data sampel

Keterangan	Jumlah Kata Awal	Jumlah Penambahan Kata	Total
Kata positif	1075	1801	2876
Kata negatif	1825	1485	3310

Tabel 11 menjelaskan jumlah penambahan kata tiap kelas sentimen yang sekaligus divalidasi pada sentimen leksikon berdasarkan data sampel. Jumlah kata tidak baku yang ditemukan pada data sampel berjumlah 1478 sehingga total jumlah data pembakuan kata sebesar 5198. Ilustrasi hasil implementasi pada model tanpa *tree* dan berbasis *tree* dapat dilihat pada Tabel 12.

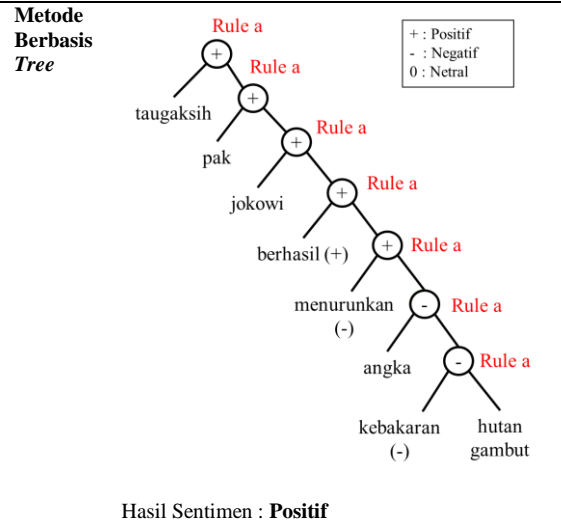
Tabel 11 Jumlah penambahan kata yang sekaligus divalidasi pada sentimen leksikon berdasarkan data sampel

Keterangan	Jumlah Kata positif	Jumlah Kata negatif
Kata Awal	1075	1825
Penambahan Kata	1808	1488
Penghapusan Kata Karena Netral	86	163
Penghapusan Kata Karena Positif	-	43
Penghapusan Kata Karena Negatif	20	-
Total	2876	3310

Contoh data *twit* pada Tabel 12 menunjukkan bahwa hasil sentimen metode berbasis *tree* sesuai dengan sentimen aktualnya, yaitu positif. Pada metode tanpa *tree*, kata "kebakaran" tidak terdapat pada sentimen leksikon sehingga kata tersebut dianggap netral. Andaikan kata "kebakaran" ada dalam sentimen leksikon dan bernilai negatif pada metode tanpa *tree*, hasil klasifikasinya tetap salah. Adapun hasil klasifikasinya adalah negatif.

Tabel 12 Hasil implementasi

Twit	RT @PutriYadis: #TauGakSih pak Jokowi berhasil menurunkan angka kebakaran hutan gambut
	Sentimen Aktual : Positif
Metode Tanpa Tree	berhasil (+), menurunkan (-). skor positif = 1, skor negatif = 1
	Hasil Sentimen : Netral



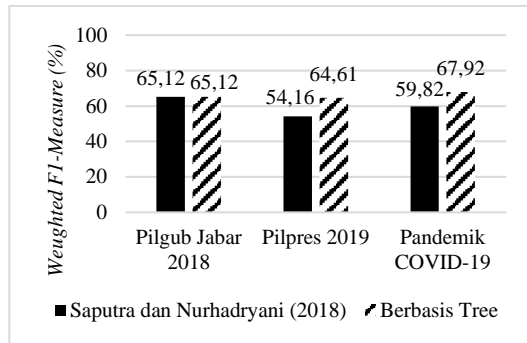
3.4. Analisis dan Evaluasi Hasil

Pengujian akurasi dilakukan dengan membandingkan kedua metode pada ketiga data. Tabel 13 menjelaskan hasil akurasi dari pengujian yang dilakukan. Tabel 13 menunjukkan bahwa metode berbasis *tree* terbukti menghasilkan akurasi yang lebih tinggi dibandingkan metode tanpa *tree* (Saputra dan Nurhadryani, 2018). Selain itu, metode berbasis *tree* menghasilkan akurasi yang lebih stabil pada seluruh data uji dengan nilai standar deviasi yang lebih kecil (0,97%).

Tabel 13 Hasil akurasi dengan membandingkan kedua metode pada ketiga data

Metode	Data Twit			Standar Deviasi
	Pilgub Jabar 2018	Pilpres 2019	Pandemik COVID-19	
Tanpa Tree	63,71%	52,92%	58,63%	5,4%
Berbasis Tree	64,97%	64,33%	66,24%	0,97%

Pengujian *weighted f1-measure* kemudian dilakukan pada kedua metode (Gambar 9) mengingat ketiga data uji memiliki komposisi kelas yang tidak seimbang. Gambar 9 menunjukkan bahwa metode berbasis *tree* dapat meningkatkan *weighted f1-measure* pada data *twit* Pilpres 2019 sebesar 13.76% dan pandemik COVID-19 sebesar 4.79%. Namun pada data *twit* Pilgub Jabar 2018 nilai *weighted f1-measure* kedua metode tidak berbeda secara signifikan. Perbandingan hasil uji akurasi dan *weighted f1-measure* dilakukan untuk mengukur validitas nilai akurasi (Tabel 14). Apabila nilai akurasi dan *weighted f1-measure* tidak berbeda secara signifikan maka nilai akurasi dianggap valid dan tidak bias terhadap data tidak seimbang.



Gambar 9 Hasil *weighted f1-measure* dengan membandingkan kedua metode pada ketiga data

Tabel 14 Perbandingan akurasi dan *weighted f1-measure*

Data	Akurasi (%)	<i>Weighted f1-measure</i> (%)	Selisih (%)
Pilgub 2018			
• Saputra dan Nurhadryani (2018)	63,71	65,12	1,41
• Berbasis <i>Tree</i>	64,97	65,12	0,15
Pilpres 2019			
• Saputra dan Nurhadryani (2018)	52,92	54,16	1,24
• Berbasis <i>Tree</i>	64,33	64,61	0,28
Pandemik COVID-19			
• Saputra dan Nurhadryani (2018)	58,63	59,82	1,19
• Berbasis <i>Tree</i>	66,24	67,92	1,68

Tabel 14 menunjukkan bahwa kedua metode memiliki selisih yang relatif kecil (kurang dari 2%) sehingga dapat dikatakan kedua nilai tersebut tidak berbeda secara signifikan. Maka dari itu dapat dikatakan bahwa akurasi kedua metode valid dan tidak bias walaupun jumlah sampel pada setiap kelas tidak seimbang.

Seluruh pengujian yang dilakukan menunjukkan bahwa metode berbasis *tree* dapat menghasilkan performa yang lebih tinggi dan stabil pada lintas topik. Hal ini menunjukkan hubungan antarkata yang direpresentasikan menggunakan struktur *tree* berpengaruh terhadap peningkatan performa analisis sentimen berbasis leksikon.

4. KESIMPULAN

Penelitian ini menunjukkan bahwa metode berbasis *tree* terbukti dapat mengatasi kelemahan pada penelitian Saputra dan Nurhadryani (2018). Metode berbasis *tree* dapat menghasilkan akurasi yang lebih tinggi dari metode tanpa *tree* (Saputra dan Nurhadryani, 2018) pada seluruh data uji. Peningkatan akurasi pada data Pilgub Jabar 2018 sebesar 1,26%, data Pilpres 2019 sebesar 11,41%, dan data Pandemi COVID-19 sebesar 7,61%.

Selain itu, metode berbasis *tree* terbukti dapat menghasilkan akurasi yang stabil pada beberapa lintas topik seperti pada data tentang pemilu dan kesehatan. Hal ini tercemin dari standar deviasi akurasi antar data uji yang kecil, yaitu sebesar 0,97%. Sedangkan metode tanpa *tree* (Saputra dan

Nurhadryani, 2018) memiliki standar deviasi akurasi yang lebih besar, yaitu sebesar 5,4%.

Pengujian *weighted f1-measure* menunjukkan bahwa pada data Pilgub 2018 kedua metode tidak berbeda signifikan. Akan tetapi pada data Pilpres 2019 metode berbasis *tree* berhasil meningkatkan *weighted f1-measure* sebesar 10,45% dan pada data pandemi COVID-19 sebesar 8,1%. Hasil akurasi dan *weighted f1-measure* memiliki selisih yang kecil pada kedua metode sehingga pengukuran akurasi kedua metode valid dan tidak bias terhadap data tidak seimbang.

Pengembangan penelitian analisis sentimen selanjutnya dapat dilakukan dengan membuat sentimen leksikon yang dapat membedakan setiap kata berdasarkan kelas katanya sehingga tidak terjadi ambiguitas. Selain itu, algoritme dalam pembentukan *tree* tidak bisa disamakan karena setiap konjungsi memiliki karakteristik masing-masing dalam pembagian klausa. Oleh karena itu, perlu kajian lebih lanjut pada seluruh konjungsi dalam bahasa Indonesia agar pembagian klausa dalam kalimat sesuai dengan kaidah bahasa Indonesia. Sebagai contoh, pada penelitian ini masih salah dalam melakukan pembagian klausa dengan konjungsi seperti “yang”. Adapun konjungsi tersebut belum melekat pada kata benda. Selain itu, *tree* yang dihasilkan belum efisien karena seluruh kata pada kalimat dibentuk dalam bentuk *tree*. Padahal belum tentu semua kata memiliki informasi yang berguna, seperti halnya pada kata keterangan. Kata keterangan hanya memberikan keterangan atau penjelasan lebih lanjut saja terhadap kata dan tidak mengubah sentimen kata. Oleh karena itu, perlu ada pemangkasan *tree* agar biaya komputasinya lebih kecil.

UCAPAN TERIMA KASIH

Penulis mengucapkan terima kasih kepada Kementerian Riset dan Teknologi / Badan Riset dan Inovasi Nasional Republik Indonesia yang telah memberikan bantuan melalui program hibah Penelitian Tesis Magister (PTM) dengan nomor 4281/IT3.L1/PN/2020 tanggal 12 Mei 2020 sehingga naskah jurnal ini dapat diterbitkan. Semoga hasil penelitian ini dapat bermanfaat bagi peneliti lain maupun masyarakat luas.

DAFTAR PUSTAKA

- ABDULLA N.A., AHMED N.A., SHEHAB M.A., dan AL-AYYOUB M., 2013. Arabic sentiment analysis: Lexicon-based and corpus-based. Dalam: *IEEE Jordan conference on applied electrical engineering and computing technologies (AEECT)*. Jordan (JO): IEEE, pp.1-6.
- ABDULLA N.A., AHMED N.A., SHEHAB M.A., AL-AYYOUB M., AL-KABI M.N., dan AL-RIFAI S., 2014. Towards improving the lexicon-based approach for arabic sentiment analysis. *International Journal of*

- Information Technology and Web Engineering, 9(3), pp.55-71.
- ALWI H., DARDJOWIDJOJO S., LAPOLIWA H., dan MOELIONO A.M., 2010. Tata Bahasa Baku Bahasa Indonesia: Edisi Ketiga. Jakarta: Pusat Bahasa dan Balai Pustaka.
- APJII (Asosiasi Penyelenggara Jasa Internet Indonesia), 2018. Penetrasi dan Perilaku Pengguna Internet Indonesia 2018, [online] Tersedia di: <<https://apjii.or.id/survei2018/download/cdKN5zPDGmQWkwvVIByZM1SlsXjaCA>> [Diakses 23 Agustus 2020].
- AZIZ A.T.A., 2013. *Sistem Pengklasifikasian Entitas pada Pesan Twitter Menggunakan Ekspresi Regular dan Naïve Bayes*. [skripsi] Institut Pertanian Bogor, Indonesia.
- BACCIANELLA S., ESULI A., dan SEBASTIANI F., 2010. Sentiwordnet 3.0: an enhanced lexical resource for sentiment analysis and opinion mining. Dalam: *Language Resources and Evaluation (LREC)*, vol. 10, no. 2010, pp. 2200-2204.
- CHEN Y., dan SKIENA S., 2014. Building Sentiment Lexicons for All Major Languages. Dalam: *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics*. Baltimore (US): ACL, pp.383-389.
- HAN J., KAMBER M., dan PEI J., 2012. Data Mining Concept and Techniques. Ed ke-3. Waltham: Elsevier Inc.
- INDURKHYA N., dan DAMERAU F.J., 2010. Handbook of Natural Language Processing Second Edition. Boca Raton: Taylor and Francis Group.
- KWAK H., LEE C., PARK H., dan MOON S., 2010. What is Twitter, a Social Network or a News Media?. Dalam: *Proceedings of the 19th International Conference on World Wide Web (WWW'10)*. Raleigh (US): ACM, pp.591-600.
- LAILIYAH M., SUMPENO S., dan PURNAMA I.E., 2017. Sentiment analysis of public complaints using lexical resources between Indonesian sentiment lexicon and Sentiwordnet. Dalam: *International Seminar on Intelligent Technology and Its Applications (ISITIA)*. Surabaya (ID). IEEE, pp. 307-312.
- LI B., ZHOU L., FENG S., dan WONG K.F., 2010. A Unified Graph Model for Sentence-Based Opinion Retrieval. Dalam: *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics (ACL)*. Uppsala (SE): ACL, hlm. 1367-1375.
- LI P., ZHU Q., dan ZHANG W., 2011. A dependency tree based approach for sentence-level sentiment classification. Dalam: *12th International Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing (ACIS)*. Sydney (AU): IEEE, pp. 166-171.
- NAKAGAWA T., INUI K., dan KUROHASHI S., 2010. Dependency tree-based sentiment classification using CRFs with hidden variables. Dalam: *Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics (ACL)*. Los Angeles (US): ACL, pp. 786-794.
- PAK A., dan PAROUBEK P., 2011. Text representation using dependency tree subgraphs for sentiment analysis. Dalam: *International Conference on Database Systems for Advanced Applications*. Berlin (DE): Springer, pp.323-332.
- PALANISAMY P., YADAV V., dan ELCHURI H., 2013. Serendio: Simple and Practical lexicon based approach to Sentiment Analysis. Dalam: *Proceedings of the Seventh International Workshop on Semantic Evaluation (SemEval)*. Georgia (GE): ACL, pp.543-548.
- PAMUNGKAS E.W., dan PUTRI D.G.P, 2016. An experimental study of lexicon-based sentiment analysis on Bahasa Indonesia. Dalam: *6th International Annual Engineering Seminar (InAES)*. Yogyakarta (ID): IEEE, pp. 28-31.
- RASHEL F., LUTHFI A., DINAKARAMANI A., dan MANURUNG R., 2014. Building an Indonesian rule-based part-of-speech tagger. Dalam: *2014 International Conference on Asian Language Processing (IALP)*. Kuching (MY): IEEE, pp.70-73.
- SAPUTRA F.T., dan NURHADRYANI Y., 2018. Analysis of Indonesian Sentiments Using Indonesian Sentiment Lexicon by Considering Denial. Dalam: *2018 International Conference on Advanced Computer Science and Information Systems (ICACSIS)*. Yogyakarta (ID): IEEE, pp.361-366.
- SIMON K., 2020. We are Social: Digital 2020 Indonesia, [online] Tersedia di: <<https://datareportal.com/reports/digital-2020-indonesia>> [Diakses 3 Oktober 2020].
- TABOADA M., BROOKE J., TOFILOSKI M., VOLL K., dan STEDE M., 2011. Lexicon-based methods for sentiment analysis. *Computational linguistics*, 37(2), pp.267-307.
- TENG Z., dan ZHANG Y., 2017. Head-lexicalized bidirectional tree lstms. *Transactions of the*

Association for Computational Linguistics,
5, pp. 163-177.

UMAR H., 2002. Metode Riset Bisnis Panduan Mahasiswa untuk Melaksanakan Riset Dilengkapi Contoh Proposal dan Hasil Riset Bidang Manajemen dan Akuntansi. Jakarta: Gramedia Pustaka Utama.

WAHYUNI E.S., KRISHANDINI, dan DEFINA, 2018. Zona Bahasa. Bogor: Idemedia Pustaka Utama.

Halaman ini sengaja dikosongkan