



Penanganan Negasi dalam Analisa Sentimen Bahasa Indonesia

Rachmad Imam Tarecha¹, Farid Wahyudi², Urnika Mudhifatul Jannah³

^{1,2,3} Fakultas Sains dan Teknologi, Universitas Islam Raden Rahmat, Indonesia

Info Artikel	ABSTRAK
Riwayat Artikel: Diterima : 30-04-2022 Direvisi : 23-05-2022 Disetujui : 31-05-2022	<p>Analisa sentimen merupakan bagian dari proses <i>opinion mining</i> yang termasuk di lingkup kerja pemrosesan bahasa alami dalam domain kecerdasan buatan. Kecerdasan buatan untuk analisa sentimen ini digunakan untuk mengotomasikan proses analisa sentimen yang biasa dilakukan manusia, namun sudah tidak memungkinkan lagi karena jumlah data yang sangat besar. Dengan teknologi pemrosesan bahasa alami, analisa sentimen dapat dilakukan secara otomatis meskipun banyak data. Teknologi analisa sentimen ini sayangnya menghadapi tantangan serius seperti adanya negasi kata atau kalimat yang membalik makna pesan. Jika negasi tidak diatasi, maka analisa sentimen menjadi tidak valid. Karena itu penelitian ini berupaya menangani negasi kata atau kalimat dalam proses analisa sentimen. Metode yang digunakan berupa proses pembalikan nilai menggunakan fungsi <i>invers</i> jika dalam sebuah kata atau kalimat memuat pesan negasi. Selain inversi juga ditambahkan mekanisme reduksi nilai untuk penanda negasi numeral tak tentu. Penggunaan inversi atau reduksi ditentukan berdasarkan klasifikasi jenis negasinya. Setelah negasi ditangani, didapati bahwa metode yang diusulkan dapat meningkatkan akurasi analisa sentimen.</p>
Kata Kunci: Sentimen Analisis, Cakupan Negasi, Penanganan Negasi,	ABSTRACT
Keywords: <i>Sentiment Analysis,</i> <i>Scope Negation,</i> <i>Negation Handling,</i>	<p><i>Sentiment analysis is part of the opinion mining process which is included in the scope of natural language processing in the artificial intelligence domain. Artificial intelligence for sentiment analysis is used to automate the sentiment analysis process that is usually done by humans, but it is no longer possible because of the very large amount of data. With natural language processing technology, sentiment analysis can be done automatically even though there is a lot of data. This sentiment analysis technology unfortunately faces serious challenges such as the negation of words or sentences that reverse the meaning of the message. If the negation is not overcome, then the sentiment analysis becomes invalid. Therefore, this study attempts to deal with the negation of words or sentences in the process of sentiment analysis. The method used is in the form of a value reversal process using an inverse function if a word or sentence contains a negation message. In addition to inversion, a value reduction mechanism is also added for the negation marker of indeterminate numerals. The use of inversion or reduction is determined based on the classification of the type of negation. After the negation is handled, it is found that the proposed method can improve the accuracy of sentiment analysis.</i></p>

Penulis Korespondensi:

Rachmad Imam Tarecha
Program Studi Teknik Informatika,
Universitas Islam Raden Rahmat
Email: ri.tarecha@gmail.com

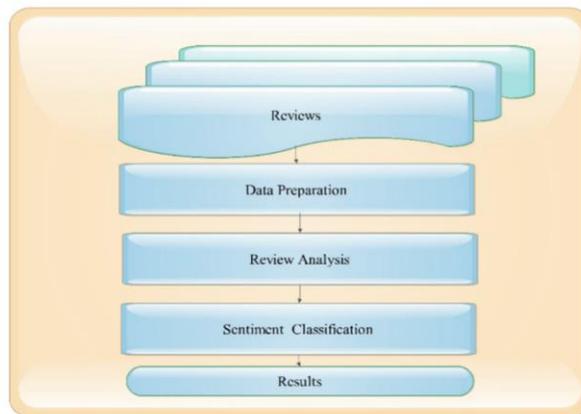
This is an open access article under the [CC BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.



1. PENDAHULUAN

Sentimen analisis merujuk pada proses untuk menentukan opini atau emosi yang dekspresikan dalam teks terhadap subjek [1]. Umar Farooq (2017) menilai bahwa sentiment analisis merupakan irisan dari pemrosesan bahasa alami, linguistik komputasional, juga *text mining* [1]. Sentimen analisis ini sendiri merupakan proses polarisasi atau klasifikasi teks ke dalam kelas positif atau negatif [2]. Selain berdasarkan teks bisa juga dengan pembobotan *emoticon* [3].

Dalam proses polarisasi tersebut terdapat gangguan atau tantangan berupa negasi yang mempengaruhi polarisasi [1]. Negasi kata itu sendiri merupakan bentuk ingkaran atau penyangkalan dari sebuah pernyataan atau kata tertentu. Jika kata "semangat" diklasifikasi menjadi kelas positif, maka seharusnya kata yang mengandung negasi seperti "tidak semangat" harus diklasifikasikan menjadi kelas negatif. Untuk itu, penting bagi proses analisa sentimen dapat menangani negasi agar terhindar dari kesalahan polarisasi.



Gambar 1 Proses Polarisasi Sentimen [4]

Proses polarisasi dalam analisa sentiment itu terdiri dari empat tahapan utama seperti pada Gambar 1 yakni akuisisi data, persiapan data atau *preprocessing*, analisa linguistik, kemudian klasifikasi sentimen. Pada tahapan pertama, data berbentuk teks diakuisisi dari beragam sumber menggunakan teknik *data mining*. Sumbernya bisa berasal dari sosial media seperti *twitter*, kolom *review* website, maupun portal berita. Setelah data teks ditambang, proses selanjutnya adalah pembersihan data yang biasanya mengandung banyak *noise*. Misalkan saja membuang data yang mengandung banyak sekali kata *out of vocabulary* atau yang tidak memiliki arti. Kemudian langkah terakhir adalah klasifikasi sentimen yang biasanya dilakukan dengan pendekatan *machine learning* [5], pendekatan berbasis leksikal, maupun gabungan dari kedua pendekatan tersebut atau *hybrid*[4].

Beberapa penelitian terkait penanganan negasi melakukan aksi penanganan negasinya pada tahapan analisis linguistik atau analisa sentiment. Penanganan negasi didahului dengan mendeteksi adanya kata yang termasuk ke dalam daftar penanda negasi.

Tabel 1 Daftar Penanda Negasi Bahasa Inggris [1]

Kelas Negasi	Penanda Negasi
Sintaksis	<i>No, not, rather, couldn't, wasn't, didn't, wouldn't, shouldn't, weren't, don't, shouldn't, weren't, don't, doesn't, haven't, hasn't, won't, wont, hadn't, never, none, nobody, nothing, neither, nor, nowhere, isn't, can't, cannot, mustn't, mightn't, shan't, without, needn't</i>
<i>Diminisher</i>	<i>Hardly, less, little, rarely, scarcely, seldom</i>
Morfologis	<i>Prefixes: de-,dis-,il-,im-,in-,ir-,mis-,non-,un-, Suffix: -less</i>

Daftar penanda negasi, dalam Bahasa Inggris dibagi menjadi tiga kelas oleh Umar Farooq (2017) menjadi daftar kata negasi berdasarkan sintaksis, berdasarkan *diminisher*, dan berdasarkan morfologis seperti pada Tabel 1 [1]. Kalimat yang memuat kata negasi dalam kelas negasi *deminisher* akan mereduksi nilai sentimen. Jika kalimat memuat kata negasi dalam kelas negasi sintaksis, maka nilai sentimen dilakukan pembalikan nilai atau inversi.

Tahapan selanjutnya setelah mendeteksi adanya kata-kata yang termasuk dalam daftar kata negasi kemudian penilaian terhadap *scope* negasi apakah itu mempengaruhi satu kata saja atau keseluruhan kalimat [1]. Umar Farooq (2017) menggunakan kelas negasi Bahasa Inggris dalam Tabel 1 untuk menentukan *scope* dari negasi apakah mempengaruhi satu kata saja atau keseluruhan kalimat [1]. Setelah negasi *scope* diketahui, barulah dilakukan pembalikan nilai atau inversi.

Sementara itu, dalam Bahasa Indonesia terdapat beberapa pendekatan untuk menangani cakupan negasi seperti *First Sentence Word* (FSW) yang menginversi nilai sentimen satu kata setelah penanda negasi. Selain itu ada juga *Fixed Window Length* (FWL) yang menginversi nilai sejumlah *length* tertentu setelah penanda negasi. Pada penelitian Annisa Maulidia Ningtyas (2018) didapati bahwa penanganan cakupan negasi dalam Bahasa Indonesia menggunakan FWL dengan *length*=3 (3 kata setelah kata negasi) memberikan performa yang baik dengan hasil akurasi sebesar 73.79% [6].

Dalam beberapa penelitian terkait penanganan negasi dalam Bahasa Indonesia, belum berdasarkan pada jenis penanda negasi, yang tidak selalu menginversi nilai namun ada juga yang mereduksi nilai. Untuk itu penelitian ini dilakukan untuk menangani kasus negasi berdasarkan jenis negasinya.

2. METODE PENELITIAN

Metode yang digunakan dalam penanganan negasi dapat dilakukan dengan membuat klasifikasi penanda negasi sesuai jenisnya, kemudian menjadikannya sebagai basis penanganan negasi. Setelah itu, hasilnya dibandingkan antara analisa sentimen tanpa penanganan negasi, dan FWL *length* 3 dengan penanda negasi konvensional, dan FWL *length* 3 dengan klasifikasi negasi yang memuat tambahan penanda negasi reduktif. Hasil yang diukur adalah perbandingan akurasi.

2.1. Penanda Negasi Konvensional dalam Bahasa Indonesia

Menurut Dian Noviani Syafar (2016) dalam Bahasa Indonesia terdapat empat kata negasi yang lazim digunakan yaitu tak atau tidak, bukan, jangan, dan belum [7]. Penggunaan kata negasi, dapat digunakan pada jenis kalimat deklaratif, interogatif, dan ekklamatif[7].

Negasi dalam Bahasa Indonesia dapat menegasikan keseluruhan kalimat, atau hanya sebagian dari kalimat. Dalam kasus tertentu, kata dengan numeral tak tentu (contohnya "sedikit") tidak membalik nilai makna secara keseluruhan. Seperti frasa dengan penanda negasi konvensional "tidak nakal" yang membalik keseluruhan nilai sentimen pada kata nakal, sementara frasa dengan numeral tak tentu "sedikit nakal" tidak membalik keseluruhan nilai sentimen, melainkan mengurangi atau mereduksi nilai sentimennya.

2.2. Daftar Penanda Negasi Bahasa Indonesia Usulan

Dengan adanya landasan teori terkait negasi atau pengingkaran dalam bahasa Indonesia, dan beberapa penelitian sebelumnya terkait penanganan cakupan negasi. Peneliti mengajukan penanganan negasi dalam analisa sentimen Bahasa Indonesia menggunakan penanda negasi yang diklasifikasikan sesuai jenis dan cakupan negasinya.

Tabel 2 Daftar Penanda Negasi Bahasa Indonesia

Kelas Negasi		Penanda Negasi Usulan
Inversi	Sebagian kalimat	tak, tidak, bukan, belum, jangan
	Seluruh kalimat	bohong
Reduksi		sedikit,

Pada Tabel 2, penanda negasi diklasifikasikan berdasarkan sifat dan cakupan negasi yakni inversi sebagian kalimat, inversi seluruh kalimat, dan reduksi. Pada kelas inversi sebagian kalimat, inversi dilakukan dengan penerapan metode FWL dengan *length* 3 kata setelah kata penanda negasi. Sementara itu kelas inversi seluruh kalimat membalikan nilai dari keseluruhan kalimat. Khusus kelas reduksi tidak dilakukan pembalikan nilai, melainkan pengurangan nilai pada nilai sentimen.

2.3. Data Leksikal

Sistem nilai dalam data leksikal yang akan digunakan dapat menggunakan data leksikal berbasis nilai seperti AFINN-111. Data leksikal AFINN-111 ini merupakan data leksikal sentimen yang bernilai antara -5 untuk kata negatif hingga +5 untuk kata positif. Penilaian pada data leksikal AFINN-111 ini dilakukan manual oleh Finn Arup Nielsen dengan isi sejumlah 2477 kata dalam Bahasa Inggris [8]. Karena dalam bentuk Bahasa Inggris, perlu diterjemahkan terlebih dahulu dalam Bahasa Indonesia .

Tabel 3 Data Leksikal AFINN-111 yang Diterjemahkan ke Bahasa Indonesia

Daftar Kata	Skor
Mengabaikan	-2
Ditinggalkan	-2
Meninggalkan	-2
Diculik	-2
Penculikan	-2
Dan seterusnya.	Dan seterusnya.

Pada Tabel 3 merupakan data leksikal AFINN-111 yang telah ditranslasikan ke Bahasa Indonesia. Pada data AFINN-111 memuat daftar kata tidak hanya kata dasarnya saja, namun dengan beragam kata imbuhan. Karena itu, AFINN-111 tidak diperlukan *pre-processing* seperti *stemming* yang mengubah kata ke bentuk dasarnya.

$$Sentiment\ Score = \frac{Total\ Scores}{Total\ Words} \tag{1}$$

Pengukuran sentiment dengan leksikal AFINN-111 dilakukan dengan menjumlahkan keseluruhan poin, dibagi dengan total kata. Hasilnya jika 0 maka tergolong netral, jika lebih dari 0 maka digolongkan sebagai positif, kemudian jika di bawah 0 digolongkan menjadi negatif.

2.4. Pseudocode

Berikut ini adalah *pseudocode* untuk analisa sentiment menggunakan AFINN-111, dengan penanganan negasi gabungan antara FWL *length* 3 dan penangananan negasi menggunakan klasifikasi yang diusulkan.

Analisa sentimen menggunakan AFINN-111, FWL dengan <i>length</i> = 3 dan klasifikasi jenis negasi	
1	procedure afinn111danfwl3danjenisnegasi(kalimat)
2	$dataleksikal \leftarrow afinn111.json$
3	$datanegasi \leftarrow datanegasi.json$
4	$skorsentimen \leftarrow declare\ an\ array$

5	<i>indexnegasi</i> ← declare an array
6	<i>skorakhir</i> ← 0
7	<i>typenegasi</i> ← declare an string
8	<i>kata</i> ← split(<i>kalimat</i>)
9	
10	For i = 0 to <i>kata</i> length – 1 do
11	<i>katalowercase</i> ← toLowerCase(<i>kata</i> [i])
12	If <i>dataleksikal</i> hasOwnProperty(<i>katalowercase</i>) then
13	<i>skorsentimen</i> push(Number(<i>dataleksikal</i> [<i>katalowercase</i>]))
14	Else
15	<i>skorsentimen</i> push(0)
16	Endif
17	If <i>datanegasi</i> hasOwnProperty(<i>katalowercase</i>) then
18	<i>indexnegasi</i> push(i)
19	<i>typenegasi</i> = <i>datanegasi</i> [<i>katalowercase</i>]
20	Endif
21	Endfor
22	
23	For j = 0 to <i>indexnegasi</i> length -1 do
24	For k = 1 to 4 do
25	If <i>skorsentimen</i> length > <i>skorsentimen</i> [j]+k then
26	If <i>typenegasi</i> is <i>sebagian</i> then
27	<i>skorsentimen</i> [<i>indexnegasi</i> [j]+k] *= -1;
28	Else If <i>typenegasi</i> is <i>reduksi</i> then
29	If <i>skorsentimen</i> [<i>indexnegasi</i> [j]+k]>0 then
30	<i>skorsentimen</i> [<i>indexnegasi</i> [j]+k] -= 1;
31	Else
32	<i>skorsentimen</i> [<i>indexnegasi</i> [j]+k] += 1;
33	Endif
34	Endif
35	Endif
36	Endfor
37	Endfor

38	
39	If <i>typenegasi</i> [<i>typenegasi.length-1</i>] Is seluruh then
40	For <i>kk</i> = 0 to <i>skorsentimen length -1</i> do
41	<i>skorsentimen</i> [<i>kk</i>] *= -1;
42	Endfor
43	Endif
44	
45	Return <i>skorsentimen</i>

Pseudocode menjelaskan bagaimana logika program berjalan dengan menganalisa terlebih dahulu sebuah teks memuat penanda negasi pada jenis apa seperti pada baris 17. Jika jenis negasi sebagian, maka menginversi nilai negasi 3 kata setelah penanda negasi. Jika jenis negasi reduksi, maka mereduksi nilai negasi 1 poin pada 3 kata setelah penanda negasi pada kode baris ke-28. Jika termasuk dalam jenis negasi keseluruhan, maka keseluruhan nilai diinversikan pada beris kode ke-39.

2.5. Dataset dan Pengukuran

Data yang digunakan untuk testing adalah data *Indonesian-Sentiment-Analysis-General-Dataset* yang berisi 10.806 baris data cuitan pengguna sosial media Twitter [9]. Rasio label atau anotasi dalam dataset tersebut adalah 2:1:1 untuk label netral, positif, dan negatif. Dataset ini dapat diunduh pada halaman <http://ugm.id/idsadataset>.

Sebagai data pembanding digunakan juga data *News Sentiment in Bahasa* yang diunggah oleh Husein Zolkepli pada halaman <https://www.kaggle.com/datasets/huseinzol05/news-sentiment-in-bahasa>. Data tersebut memiliki rasio label positif 55%, negatif 26%, dan sisanya 19% dari total 3834 data.

Pengukuran yang digunakan dalam penelitian ini menggunakan *confusion* matrix dengan dimensi 3X3 untuk variable ukur positif, negatif, dan netral. Pengukuran dengan *confusion* matrix dapat mengetahui tingkat akurasi, presisi, *recall*, dan *F-Score* [10].

$$Akurasi = \frac{True\ Positif\ (TP) + True\ Negatif\ (TN) + True\ Netral\ (TNET)}{TP + False\ Positif\ (FP) + TN + False\ Negatif\ (FN) + TNET + False\ Netral\ (FNET)} \tag{2}$$

$$Presisi = \frac{TP}{TP + FP} \tag{3}$$

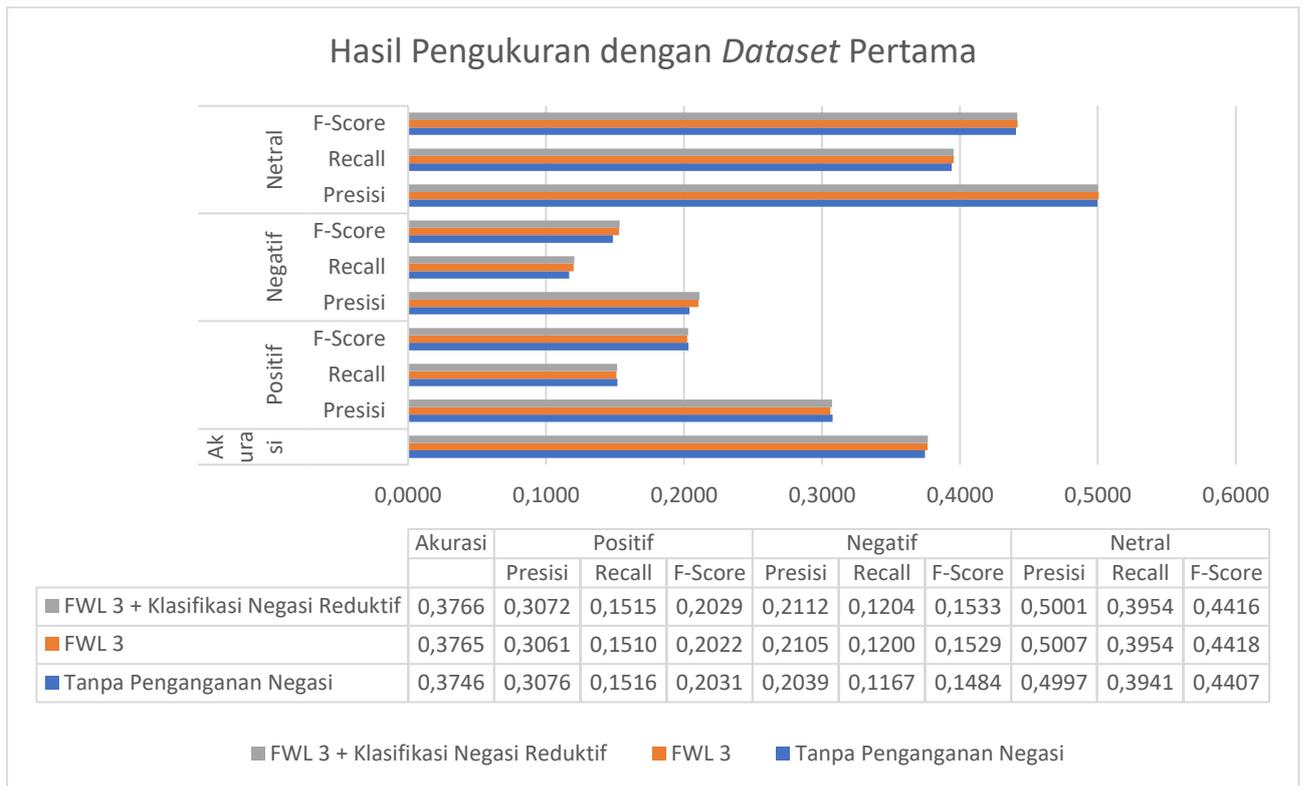
$$Recall = \frac{TP}{TP + FN + FNET} \tag{4}$$

$$Nilai - F = \frac{2 * Presisi * Recall}{Presisi + Recall} \tag{5}$$

Pengukuran akurasi merupakan pengukuran kesesuaian atau kedekatan nilai prediksi analisa sentimen dengan label aktual yang ada dalam dataset. Pengukuran presisi, *recall*, dan Nilai-F (*F-Score*) dihitung setiap label sentimenya seperti positif, negatif, dan netral.

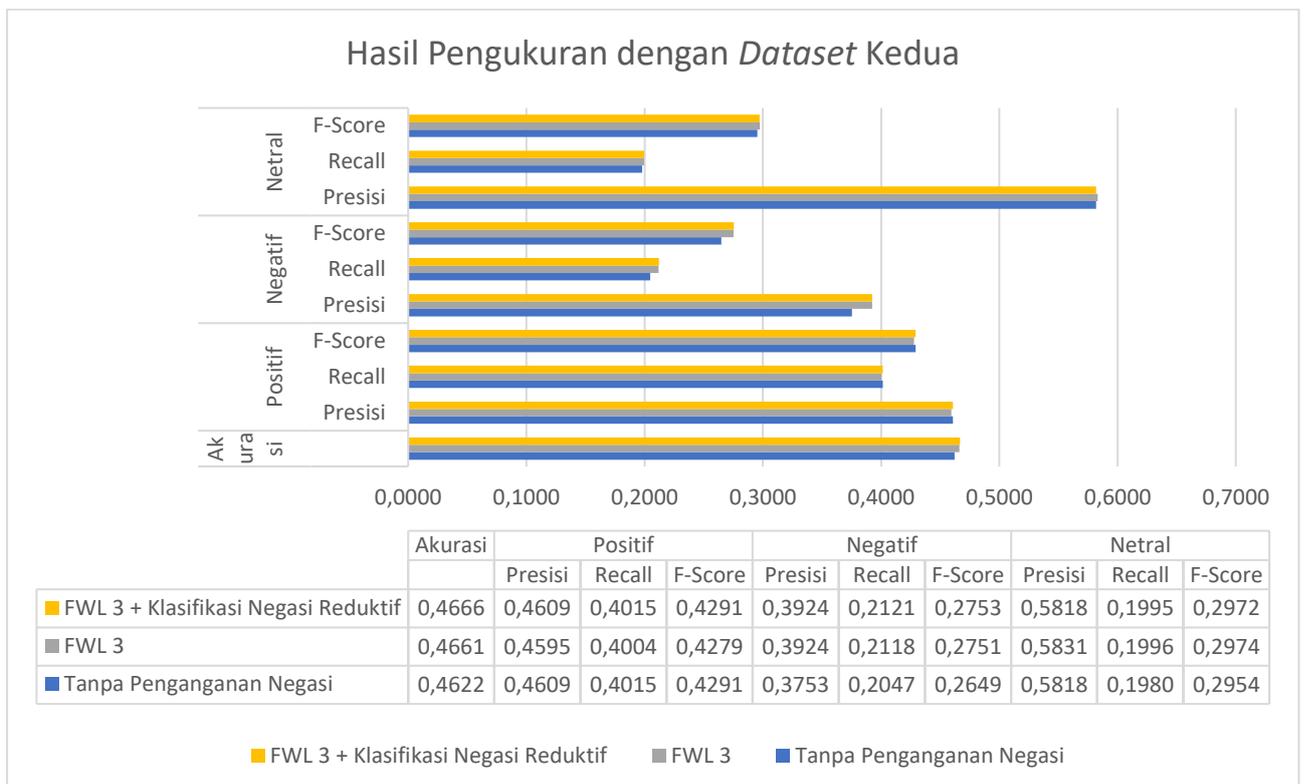
3. HASIL DAN ANALISIS

Pengukuran pertama menggunakan dataset *Indonesian-Sentiment-Analysis-General-Dataset* yang memiliki 10.806 jumlah data. Hasilnya dicatat perolehan nilai untuk akurasi, presisi, *recall*, dan *F-Score*. Pencatatan kemudian disajikan dalam bentuk *chart* untuk membandingkan nilainya.



Gambar 2 Hasil Pengukuran dengan *Dataset* Pertama

Hasil pengukuran pada Gambar 2 menunjukkan hasil akurasi 37.46% pada analisa sentimen tanpa penanganan negasi. Sementara itu penanganan negasi menggunakan metode FWL dengan *length window* 3 kata setelah pananda negasi menghasilkan nilai akurasi sebesar 37.65%. Metode yang diusulkan, FWL *length window* 3 dengan tambahan penanda negasi berupa kata numeral tak tentu seperti "sedikit" yang tidak menginversi nilai melainkan mereduksi nilai menghasilkan nilai akurasi sebesar 37.66%.



Gambar 3 Hasil Pengukuran dengan *Dataset* Kedua

Hasil pengukuran pada Gambar 3 tak jauh berbeda dengan Gambar 2. Hasil akurasinya sebesar 46.22% untuk analisa sentimen tanpa penanganan negasi. Meningkatkan menjadi 46.61% dengan penanganan negasi FWL 3. Kemudian meningkat menjadi 46.66% dengan penambahan kelas negasi reduktif yang diusulkan.

4. KESIMPULAN

Penambahan kata numeral tak tentu seperti "sedikit" ke dalam daftar dan kelas negasi reduktif yang tidak menginversi nilai melainkan mereduksi nilai, dapat menambah tingkat akurasi. Pada *dataset* pertama ada peningkatan sebesar 0.0185% pada nilai akurasi menggunakan penambahan kelas negasi reduktif, jika dibandingkan dengan FWL 3 tanpa klasifikasi negasi reduktif. Pada *dataset* kedua, tingkat akurasi meningkat sebesar 0.05% menggunakan klasifikasi penanda negasi reduktif, dibandingkan dengan FWL 3 tanpa klasifikasi negasi reduktif.

Meskipun terjadi peningkatan akurasi, namun peningkatannya tidak begitu signifikan dengan peningkatan di bawah 1%. Penelitian lanjutan yang diperlukan adalah mencari metode yang tepat dalam tahapan reduksi nilainya, untuk meningkatkan akurasi. Selain itu, penelitian lanjutan juga dapat menambahkan daftar penanda negasi lainnya yang membalik nilai sebagian, membalik nilai keseluruhan, atau mereduksi nilai.

UCAPAN TERIMAKASIH

Terima kasih kepada semua pihak yang terlibat dalam mensukseskan penelitian ini.

REFERENSI

- [1] U. Farooq, "Negation Handling in Sentiment Analysis at Sentence Level," *Journal of Computers*, pp. 470–478, 2017, doi: 10.17706/jcp.12.5.470-478.
- [2] S. Almatarneh and P. Gamallo, "Automatic Construction of Domain-Specific Sentiment Lexicon for Polarity Classification," in *The PAAMS Collection - 15th International Conference, 2017*, vol. 619. doi: 10.1007/978-3-319-61578-3.
- [3] A. Rossi, T. Lestari, R. Setya Perdana, and M. A. Fauzi, "Analisis Sentimen Tentang Opini Pilkada Dki 2017 Pada Dokumen Twitter Berbahasa Indonesia Menggunakan Naïve Bayes dan Pembobotan Emoji," 2017. [Online]. Available: <http://j-ptiik.ub.ac.id>
- [4] K. Mite-Baidal, C. Delgado-Vera, E. Solís-Avilés, A. H. Espinoza, J. Ortiz-Zambrano, and E. Varela-Tapia, "Sentiment analysis in education domain: A systematic literature review," in *Communications in Computer and Information Science*, 2018, vol. 883, pp. 285–297. doi: 10.1007/978-3-030-00940-3_21.
- [5] C. Dhaoui, C. M. Webster, and L. P. Tan, "Social media sentiment analysis: lexicon versus machine learning," *Journal of Consumer Marketing*, vol. 34, no. 6, pp. 480–488, 2017, doi: 10.1108/JCM-03-2017-2141.
- [6] A. M. Ningtyas and G. B. Herwanto, "The Influence of Negation Handling on Sentiment Analysis in Bahasa Indonesia," 2018.
- [7] D. N. Syafar, "NEGASI DALAM BAHASA INDONESIA DAN BAHASA INGGRIS," *Jurnal Arbiter*, vol. 3, no. 1, 2016.
- [8] M. A. Al-Shabi, "Evaluating the performance of the most important Lexicons used to Sentiment analysis and opinions Mining," *International Journal of Computer Science and Network Security*, vol. 20, no. 1, 2020, [Online]. Available: <https://www.researchgate.net/publication/343473213>
- [9] R. Ferdiana, F. Jatmiko, D. D. Purwanti, A. Sekar, T. Ayu, and W. F. Dicka, "Dataset Indonesia untuk Analisis Sentimen," 2019.
- [10] O. Caelen, "A Bayesian interpretation of the confusion matrix," *Annals of Mathematics and Artificial Intelligence*, vol. 81, no. 3–4, pp. 429–450, Dec. 2017, doi: 10.1007/s10472-017-9564-8.