

Analisis Sentimen Konten Radikal dalam Kontestasi Politik 2019 di Media Twitter Menggunakan Interjection dan Punctuation

Taufik Nurahman*¹, Yufis Azhar², Nur Hayatin³

^{1,2,3}Teknik Informatika/Universitas Muhammadiyah Malang

tnurahman@gmail.com*¹, yufis.az@gmail.com², noorhayatin@gmail.com³

Abstrak

Analisis sentiment saat ini menjadi tren untuk mengidentifikasi opini serta emosi seseorang dalam menanggapi sebuah situasi. Di tahun politik, sangat banyak opini bertebaran baik yang dituliskan di media cetak maupun media sosial. Para pelaku politik memiliki pandangan yang berbeda-beda, sehingga memunculkan banyak opini yang berujung pada tindakan radikal seperti perlakuan SARA kepada orang yang berbeda pandangan. Penelitian terkait dengan analisis sentimen radikal melalui media twitter sudah dilakukan oleh beberapa peneliti sebelumnya, namun belum ada penelitian analisis sentimen radikal yang menggunakan fitur ekstraksi. Penelitian ini mengusulkan untuk melakukan analisis sentimen konten radikal pada tweet tekstual berbahasa Indonesia terkait dengan kontestasi politik di Indonesia yang lalu menggunakan dua fitur yakni punctuation dan interjection, serta diklasifikasikan menggunakan algoritma support vector machine. Dari hasil klasifikasi yang telah dilakukan, didapatkan nilai akurasi dari analisis sentimen sebesar 80% dan dilakukan analisis sentiment radikal beberapa kali dengan jumlah interjection yang berbeda, didapatkan hasil akurasi sebesar 94% dengan menggunakan 200 kata interjection.

Kata Kunci: *Interjection, Punctuation, Support Vector Machine, Analisis Sentimen, Radikal*

Sentiment analysis is now a trend to identify people's opinions and emotions in responding to a situation. In the political year, many opinions were scattered both written in print and social media. Political actors have different views, so that raises a lot of opinions that lead to radical actions such as SARA to people with different opinions. Research related to the analysis of radical sentiments via Twitter has been done by several researchers before, but there have been no studies of radical sentiment analysis using extraction features. This study proposes to conduct a radical content sentiment analysis on Indonesian textual tweets related to political contestation in Indonesia which then uses two features namely punctuation and interjection, and is classified using the support vector machine algorithm. From the results of the classification that has been done, obtained an accuracy value of 80% sentiment analysis and radical sentiment analysis conducted several times with a number of different interjection, obtained an accuracy of 94% using 200 interjection words.

Keywords: *Interjection, Punctuation, Support Vector Machine, Sentiment Analysis, Radical*

1. Pendahuluan

Twitter merupakan salah satu media sosial yang memiliki jumlah basis pengguna yang sangat besar. Pertumbuhan pengguna Twitter sangatlah pesat hingga bulan Januari 2013 diketahui jumlah pengguna terdaftar mencapai 500 juta akun di seluruh dunia [1]. Bahkan, dalam penelitian yang dilakukan oleh Ghulam Asrofi Buntoro pada tahun 2016 dan diketahui jumlah pengguna yang terdaftar sebanyak 106 juta akun di Indonesia dan terus bertambah sebanyak 300.000 setiap harinya [2]. Salah satu negara yang memiliki basis pengguna terbesar Twitter adalah Indonesia, dimana penggunaannya cenderung hyper aktif dalam platform tersebut.

Penggunaan platform media social Twitter yang cenderung mudah dipakai oleh berbagai kalangan usia menjadikan Twitter sebagai salah satu platform utama dalam bertukar pendapat, menceritakan keseharian, sampai pada mengemukakan pendapat atau opini atas suatu isu. Kecenderungan pemakaian Twitter meningkat dalam segi kuantitas tweet (postingan di dalam twitter) terutama pada masa tahun-tahun politik di Indonesia, dimana opini publik oleh berbagai kalangan masyarakat meramaikan platform tersebut. Aktivitas melakukan posting berbagai tweet

dalam isu-isu politik menjadi hal yang bersifat adiktif. Hal tersebut dalam beberapa kurun waktu terakhir, menjadi penyebab terciptanya hashtags ikonik dari beberapa golongan masyarakat yang hingga saat ini digunakan sebagai simbol atau representasi dari suatu ide oleh kelompok tertentu di dalam Twitter.

Hashtag ikonik dalam platform Twitter menjadi perwakilan atas sebuah ide yang dicanangkan terhadap isu tertentu. Beberapa hashtag populer dan yang paling menonjol beberapa waktu belakangan ini seperti #2019GantiPresiden dan #2019TetapJokowi dengan berbagai macam isi tweet di dalamnya. Hashtags tersebut tentunya berpengaruh terhadap arah berpendapat dari para pengguna twitter di Indonesia saat ini.

Pengaruh di dalam hashtags yang bermuatan politik memiliki efek terhadap penggiringan opini publik dimana terdapat dua jenis yakni menggiring kepada arah positif dan juga ke arah negative seperti munculnya Hatespeech serta tweet yang mengandung opini radikalisme dari sebuah kelompok untuk mempengaruhi persepsi pengguna lain yang sebenarnya tidak terlibat di dalamnya. Kemunculan beberapa hashtags ikonik bermuatan politik juga diprakarsai oleh beberapa tweet yang juga di akomodir serta di follow up oleh tokoh-tokoh utama dari pelaku politik di Indonesia. Kuantitas tweet yang berkaitan dengan hashtag ikonik bermuatan politik, didukung dengan adanya trend yang memunculkan tweet-tweet tersebut sebagai wajah dari Twitter pada kurun waktu tertentu menjadikan semakin banyak orang untuk bergabung dan saling lempar opini atas nama hashtag tersebut. Kondisi ini berujung pada munculnya sifat sentimental dari beberapa orang yang bersifat ekstrem atau militan terhadap pihak yang mewakili nilai-nilai yang berlawanan dengan ide yang direpresentasikan dalam hashtag ikonik, misalnya calon presiden dan wakil presiden yang berlawanan dari penggagas hashtag. Situasi ini berujung pada terciptanya suasana yang tidak kondusif sehingga lagilagi berujung pada terciptanya kata umpatan untuk suatu kelompok seperti 'Cebong' dan 'Kampret'.

Pasca digunakannya kedua istilah umpatan itu dalam setiap opini yang di publikasikan melalui media twitter secara masif, kemudian terjadilah penggiringan opini yang berakibat pada opini berpolitik yang dimbumbui dengan aroma radikalisme dan ekstrimisme atas dasar sentimental satu sama lain yang mulai terjadi akhir-akhir ini. Isu-isu Suku, Agama, Ras, dan Antargolongan (selanjutnya disebut SARA) mulai dimunculkan lagi dan yang paling menonjol ialah politisasi agama yang didalamnya terdapat penggiringan pengkafiran terhadap sesama muslim (selanjutnya disebut takfiri) serta penggiringan opini kepada tindak anarkis. Hal tersebut tentu ditakutkan mengurangi esensi publik terhadap pesta demokrasi itu sendiri dan akan berakibat fatal dengan kurangnya kepercayaan masyarakat pada proses pemilihan umum (selanjutnya disebut pemilu) yang akan berlangsung.

Beberapa penelitian mengenai analisis sentiment telah menunjukkan hasil yang berbeda seperti pada penelitian ini [3] menganalisa sentiment konten radikal terorisme dari media twitter dengan menggunakan tweet berbahasa Indonesia. Tweet radikal tersebut diklasifikasikan menjadi 2 yakni radikal negative dan radikal positif namun tingkat akurasi yang didapatkan masih sebesar 70%. Namun pada penelitian [4] membahas tentang deteksi sarkas pada tweet berbahasa Indonesia menggunakan dua buah feature extraction yakni punctuation dan interjection lalu mengklasifikasikannya menggunakan naïve bayes classifier dan k-Nearest Neighbor. Hasil akhir yang didapatkan mesin hanya mampu mengklasifikasikan sedikit tweet positif sarkasme dari dua kelas yang diberikan menggunakan naïve bayes. Dibuktikan dengan nilai Akurasi sebesar 70%. Namun, menggunakan kNearest Neighbor akurasi yang didapatkan sudah cukup baik yakni sebesar 82%.

Peneliti kemudian tertarik untuk melakukan penelitian atas isu penggunaan platform Twitter yang memuat konten radikal dalam kontestasi politik 2019 dengan menganalisis sentiment yang termuat di dalamnya. Penelitian Peneliti didasarkan pada beberapa penelitian terdahulu di atas, namun dalam penelitian ini diusulkan sentiment analisis terhadap konten radikal pada tweet berbahasa Indonesia terkait dengan kontestasi politik di Indonesia dengan menggunakan metode Support Vector Machine yang ditambahkan Fitur Ekstraksi Interjection dan Punctuation sebagai fitur untuk mendeteksi tweet dengan maksud radikal. Titik perbedaan penelitian ini dengan sebelumnya ialah penggunaan metode Support Vector Machine di tambahkan dengan fitur ekstraksi Interjection dan Punctuation yang diharapkan dapat meningkatkan akurasi.

2. Metode Penelitian

2.1 Analisa Dataset

Data yang akan digunakan pada penelitian ini merupakan tweet berbahasa Indonesia yang terkait dengan kontestasi politik Indonesia pada tahun 2019. Data yang digunakan sebanyak dua ribu tweet. Yang mana tweet tersebut mengandung keywords: #2019GantiPresiden, #Jokowi2Periode, Prabowo, Jokowi, Antek Aseng, dan Politisasi Agama. Tweet yang dimaksud memiliki hubungan dengan makna dari radikal [5]. Data tweet yang berjumlah dua ribu tersebut akan di pecah kedalam bentuk data latih dan data uji, dengan rasio 80% data latih dan 20% data uji, serta rasio pembagian antara data negatif dan positif masing-masing 50% dari total data.

2.2 Lingkungan Kerja

Penelitian ini tentunya membutuhkan Software dan Hardware guna menunjang proses analisis sentimen. Hal ini diperlukan untuk menentukan pendapat dan subjektivitas dalam sebuah teks sebagai bagian dari analisis sentiment [6]. Pada penelitian ini menggunakan *software* yang ditunjukkan pada Tabel 1 berikut.

Tabel 1. Perangkat Lunak

No	Spesifikasi
1	Python versi 3.7
2	Jupyter Notebook
3	Library NLTK versi 3.4
4	Library Scikit_learn versi 0.21.3
5	Library Sastrawi versi 1.2.0

Hardware yang digunakan pada penelitian ini, ialah menggunakan sebuah laptop merk ASUS ROG tipe G550JX dengan spesifikasi perangkat keras yang ditunjukkan pada Tabel 2 berikut.


Tabel 2. Perangkat Keras

No	Spesifikasi
1	Processor Intel Core i7-4720HQ CPU @ 2.60Ghz
2	RAM 8 GB DDR3
3	Harddisk 1 TB

2.3 Pengumpulan Data atau *Crawling*

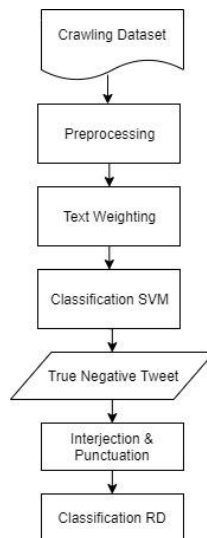
Data yang akan digunakan akan didapatkan melalui media sosial Twitter berdasarkan kriteria yang telah di jelaskan pada bagian Analisa Dataset. Seperti pada Tabel 3, *crawling* dilakukan dengan menggunakan tool TwitterScrapper. Untuk menggunakan tool TwitterScrapper ini sangat mudah. Namun, diperlukan instalasi *library* TwitterScrapper terlebih dahulu. Jika telah dilakukan instalasi, proses *Crawling* dapat dilakukan dalam Command Prompt dengan mengetikkan *syntax* dari tool TwitterScrapper:

Tabel 3. Contoh Data yang telah di *Crawling*

No	Tweet
1.	Emang loe kira Prabowo Pribumi? Kagak liat kakaknya Cina juga? Bapak nya Cina ibunya cina? Terus Prabowo Pribumi sendiri? Brojol dari mana? 
2.	Rakyat sudah tahu jok, Jokowi (01) didukung kelompok politisi preman, Syiah, liberal, munafikun dan pemuja PKI. Pak Prabowo (02) didukung kelompok religius lurus dan politisi santun serta rakyat cerdas yang waras. Faham antum??? #01vsrakyatreligiusrurus #02vsPKI

2.4. Analisa dan Perancangan

Tahapan yang akan dilakukan untuk melakukan analisis sentiment dilanjutkan dengan deteksi sarkasme pada penelitian ini dapat dilihat pada Gambar 1.



Gambar 1. Alur Proses Analisa dan Perancangan

Pada Gambar 1 di atas merupakan alur proses Analisa dan perancangan yang mana bagian-bagian dari alur tersebut akan di jelaskan pada sub bab di bawah ini.

2.4.1 Preprocessing

Untuk melakukan *preprocessing* tahapan yang dilakukan adalah *case folding*, *filtering*, dan *stemming* [7].

Terdapat 4 proses dalam melakukan *preprocessing* yaitu :

a. Labelling

Pada tahap ini, data yang telah di *crawling* akan di beri label terlebih dahulu. Data akan di labeli sebagai *tweet* yang positif atau *tweet* yang negatif secara manual.

b. Case Folding

Pada tahap *Case Folding*, huruf kapital dan non kapital pada semua *tweet* yang telah di *crawling* akan di setarakan. Contoh hasilnya dapat dilihat pada Tabel 4 di bawah.

Tabel 4. Case Folding pada Dataset

Tahap	Tweet
Awal	Emang loe kira Prabowo Pribumi? Kagak liat kakaknya Cina juga? Bapak nya Cina ibunya cina? Terus Prabowo Pribumi sendiri? Brojol dari mana?
Hasil <i>case folding</i>	emang loe kira prabowo pribumi? kagak liat kakaknya cina juga? bapak nya cina ibunya cina? terus prabowo pribumi sendiri? brojol dari mana?

Pada Tabel 4 di atas, perbedaan dari hasil *case folding* dengan data awal telah terlihat dengan kata "Emang" yang memiliki huruf kapital, menjadi "emang" yang mana huruf kapital tadi telah di ubah ke bentuk huruf kecil.

c. Filtering

Data yang telah melewati *case folding* tentunya masih menyisakan hal seperti Username, URL, Hashtag, tanda baca, ataupun Retweet. Pada tahap ini akan dilakukan penghapusan pada bagian-bagian tersebut seperti yang di ditampilkan pada Tabel 5 di bawah ini.

Tabel 5. *Filtering pada Dataset*

Tahap	Tweet
Awal	emang loe kira prabowo pribumi? kagak liat kakaknya cina juga? bapak nya cina ibunya cina? terus prabowo pribumi sendiri? brojol dari mana?
Hasil <i>filtering</i>	emang loe kira prabowo pribumi kagak liat kakaknya cina juga bapak nya cina ibunya cina terus prabowo pribumi sendiri brojol dari mana

Pada tabel di atas, dapat dilihat bahwa tanda tanya (?) yang ada pada data awal di hapus, sehingga tidak terdapat lagi pada data pasca melewati tahap *filtering*.

d. *Tokenizing*

Tahap ini mengubah data yang sebelumnya berbentuk kalimat menjadi potongan-potongan per-kata seperti contoh pada Tabel 6.

Tabel 6. *Tokenizing pada Dataset*

Tahap	Tweet
Awal	emang loe kira prabowo pribumi kagak liat kakaknya cina juga bapak nya cina ibunya cina terus prabowo pribumi sendiri brojol dari mana
Hasil Stopword Removal	'emang', 'loe', 'prabowo', 'pribumi', 'kagak', 'liat', 'kakak', 'cina', 'nya', 'cina', 'cina', 'prabowo', 'pribumi', 'brojol'.

Pada tabel diatas, terlihat perbedaan data awal dengan data yang telah melewati fase *tokenizing*. Data yang sebelumnya berbentuk kalimat, berubah menjadi potongan per-kata.

e. *Stopword Removal*

Tahap ini merupakan tahap penghapusan pada kata-kata yang tidak di butuhkan. Contoh hasilnya dapat dilihat pada Tabel 7.

Tabel 7. *Stopword Removal pada Dataset*

Tahap	Tweet
Awal	emang loe kira prabowo pribumi kagak liat kakaknya cina juga bapak nya cina ibunya cina terus prabowo pribumi sendiri brojol dari mana
Hasil Stopword Removal	Emang loe prabowo pribumi kagak liat kakak cina nya cina cina prabowo pribumi brojol

Pada tabel diatas, terlihat perbedaan data awal dengan data yang telah di *Stopword Removal*. Terutama pada kata-kata yang tidak dibutuhkan, seperti "nya", "dari", dan sebagainya.

f. *Stemming*

Tahap ini merupakan tahap pencarian kata dasar pada kata yang berimbuhan. Contoh hasilnya dapat dilihat pada Tabel 8 berikut ini.

Tabel 8. *Stemming pada Dataset*

Tahap	Tweet
Awal	Memang ada carut marut dlm segala pengelolaan di negara ini termasuk masalah kesehatan dan BPJS. Jika tidak mampu...serahkan pd yg mampu. InsyaAllah: #2019GantiPresiden #IndonesiaMenang #2019PrabowoPresidenRI

Hasil <i>stemming</i>	memang ada carut marut dlm segala kelola di negara ini masuk masalah sehat dan bpjs jika tidak mampu serah pd yg mampu insyaallah gantipresiden indonesia menang prabowo presiden ri
-----------------------	--

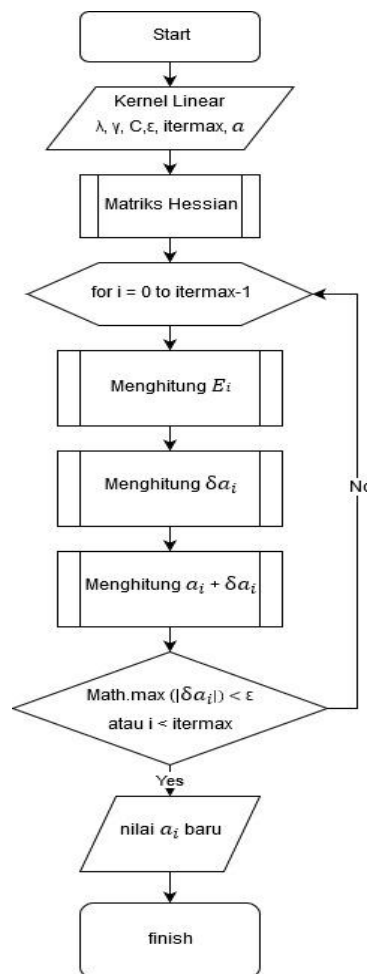
Pada tabel diatas, kata berimbuhan akan di ubah menjadi bentuk dasarnya. Terdapat beberapa perbedaan dari data awal dengan data yang telah di *Stemming*. Contohnya seperti kata “pengelolaan” yang berubah menjadi “kelola” setelah di *Stemming*.

2.4.2 Text Weighting

Pada penelitian ini, data yang telah terstruktur pasca melewati fase *preprocessing* akan di beri perhitungan bobot dari masing-masing *term* menggunakan metode *Bag of Words* dari *library ScikitLearn*. Namun, langkah pertama sebelum di lakukannya penghitungan *term*, ialah melakukan pembagian antara data *train* dan data *test*. Data yang dibubuhi fitur ini akan diubah menjadi list string, dengan prinsip *bag of words* menghindari pengulangan kata. Yang mana nantinya akan membentuk *feature vector* yang disebut sebagai dimensi.

2.4.3 Classification SVM

Setelah melewati fase *Preprocessing* dan *text weighting* atau pembobotan kata. Data akan siap untuk di klasifikasikan, guna mendapatkan nilai *true negative*. Data akan di klasifikasikan menggunakan metode *Support Vector Machine*. Metode ini akan mencari sebuah hyperplane atau pembatas pada tiap-tiap antar kelas data yang memiliki margin atau jarak hyperplane pada data paling dekat dalam setiap kelasnya [8]. Algoritma dari metode *Support Vector Machine* yang akan digunakan adalah sekuensial dengan alur *flowchart* seperti Gambar 2 berikut [9].



Gambar 2. Algoritma Sekuensial Support Vector Machine

Penjelasan langkah-langkah Sekuensial dari Gambar 2 sebagai berikut[10] :

1. Menginisialisasi parameter λ (variable scalar), γ (Learning Rate), C (variable Slack), ϵ (epsilon) dan iteraksi maksimum
2. Menginisialisasi $a_i = 0$. kemudian menghitung Matrik Hessian berikut $D_{ij} = y_i y_j (K(x_i, x_j) + \lambda^2)$
3. Melakukan tiga langkah dibawah untuk seluruh data yang ada untuk memperbaharui nilai E (error) dan a (alpha).
 - a. $E_i = \sum_{j=1}^n a_j D_{ij}$
 - b. $\delta a_i = \min \{ \max[\gamma(1 - E_i), -a_i], C - a_i \}$
 - c. $a_i = a_i + \delta a_i$
4. Kembali ke langkah 3 sampai a konvergen
5. Selesai.

2.4.4 Interjection dan Punctuation

Data telah di ketahui nilainya pasca melewati proses klasifikasi dengan metode *Support Vector Machine*, kemudian akan di ambil yang bernilai *true negative*. Data yang bernilai *true negative* tersebut akan di berikan fitur ekstraksi, yakni *Interjection* dan *Punctuation*. Data yang telah di tambahkan fitur tersebut akan di klasifikasikan kembali.

Karena saat *preprocessing* semua tanda baca akan di hapus, maka untuk dapat mengimplementasikan pemberian fitur *punctuation*, penulis kembali menggunakan data awal yang di sesuaikan dengan data hasil klasifikasi, agar dapat mengembalikan nilai 1 jika menemukan *punctuation* di dalam sebuah *tweet*.

2.5 Pengujian

Confusion Matrix merupakan metode pengujian yang digunakan untuk mengetahui tingkat keakuratan suatu model teknik klasifikasi. Terdapat beberapa jenis pengujian yang akan di lakukan di metode pengujian confusion matrix ini, antara lain *recall*, *precision*, *F1-Score*, dan *accuracy*. Berikut Tabel 9 confusion matrix.

Tabel 9. Tabel Confusion Matrix

		Actual	
		Yes	No
Prediksi	Yes	TP	FP
	No	FN	TN

Yang mana:

- a. TP (True Positive) = Jumlah sampel bernilai true yang diprediksi benar
- b. TN (True Negative) = Jumlah sampel bernilai false yang diprediksi secara benar
- c. FP (False Positive) = Jumlah sampel bernilai false yang salah diprediksi sebagai sampel bernilai true
- d. FN (False Negative) = Jumlah sampel bernilai true yang salah diprediksi sebagai sampel bernilai true

3. Hasil Penelitian dan Pembahasan

3.1 Hasil Analisis Sentimen

Pada tahap ini, di lakukan perhitungan pengujian dari hasil klasifikasi sentimen analisis yang telah dilakukan menggunakan algoritma *support vector machine*, yang mana perhitungan tersebut akan mengeluarkan nilai seperti akurasi, *recall*, dan *precision*. Berikut Tabel 10l hasil dari pengujian.

Tabel 10. Hasil pengujian sentiment analysis

	Precision	Recall	F1 Score	Akurasi
Positif	90%	72%	80%	80%
Negatif	71%	90%	79%	

Pada Tabel 10 di atas, dapat di ketahui nilai-nilai yang keluar dari masing-masing tipe pengujian. pNilai akurasi yang di dapatkan dari pengujian analisis sentimen ini sebesar 80%.

3.2 Hasil Sentimen Radikal

Sama seperti tahap pengujian sentimen analisis sebelumnya, di lakukan perhitungan pengujian dari hasil klasifikasi sentimen radikal yang beberapa kali telah dilakukan menggunakan algoritma SVM, yang mana perhitungan tersebut akan mengeluarkan nilai seperti akurasi, *recall*, dan *precision*. Berikut tabel hasil dari pengujian.

Tabel 11. Hasil Pengujian Sentiment Radikal tanpa Fitur Punctuation dan Interjection

	Precision	Recall	F1 Score	Akurasi
Non Radikal	91%	91%	93%	93%
Radikal	95%	95%	93%	

Pada Tabel 11 di atas merupakan hasil klasifikasi dari data berkategori true negative yang di dapatkan dari hasil klasifikasi analisis sentiment positif dan negatif. Klasifikasi dilakukan tanpa menambahkan kedua fitur. Nilai akurasi yang didapatkan pada pengujian sentimen radikal ini ialah sebesar 93%.

Tabel 12. Hasil Pengujian Sentimen Radikal dengan Fitur Punctuation dan Interjection n=200

	Precision	Recall	F1 Score	Akurasi
Non Radikal	92%	97%	94%	94%
Radikal	97%	92%	94%	

Tabel 12 merupakan hasil klasifikasi dari data berkategori true negative yang di dapatkan dari hasil klasifikasi analisis sentiment positif dan negatif. data tersebut kemudian di tambahkan dengan fitur punctuation dan interjection dengan jumlah 200 kata interjection. Nilai akurasi yang di dapatkan pada pengujian sentimen radikal ini meningkat menjadi 94%.

Tabel 13. Hasil Pengujian Sentimen Radikal dengan Fitur Punctuation dan Interjection n=150

	Precision	Recall	F1 Score	Akurasi
Non Radikal	91%	91%	93%	93%
Radikal	96%	96%	93%	

Pada Tabel 13 di atas merupakan hasil klasifikasi sentiment radikal dengan data true negative, masih dengan cara yang sama namun dengan jumlah interjection yang berbeda yakni sebanyak 150 kata interjection. Nilai akurasi yang di dapatkan menjadi 93%.

Tabel 14. Hasil Pengujian Sentimen Radikal dengan Fitur Punctuation dan Interjection n=100

	Precision	Recall	F1 Score	Akurasi
Non Radikal	91%	97%	94%	94%
Radikal	97%	91%	94%	

Pada Tabel 14 di atas terlihat tingkat akurasi meningkat lagi dengan menggunakan interjection sebanyak 100 kata. Sama halnya dengan percobaan pada jumlah interjection sebanyak 200 kata yang memiliki tingkat akurasi 94%, namun keduanya memiliki perbedaan tipis pada nilai precision dan recall.

4. Kesimpulan

Tahun politik merupakan tahun yang rawan akan tulisan-tulisan yang bernada negatif, terutama tulisan-tulisan para pelaku politik di twitter. Penelitian ini mengusulkan untuk melakukan analisis sentimen guna mengambil data yang bernilai negatif yang termasuk kedalam kategori true negative dan di lakukan penambahan fitur kedalam data tersebut.

Terdapat dua fitur yang di gunakan pada tweet yang bernilai true negative, yakni punctuation dan interjection. Data yang telah di berikan fitur tersebut, akan di klasifikasikan kembali dengan algoritma SVM. Berdasarkan hasil penelitian, tingkat akurasi yang di peroleh pada dari analisis sentimen yang diklasifikasikan dengan algoritma SVM yang di sediakan oleh library scikit-learn menunjukkan angka 80%. Data true negative yang didapatkan dari hasil klasifikasi analisis sentimen sebelumnya akan diklasifikasikan kembali menggunakan algoritma support vector machine dengan dan tanpa fitur punctuation dan interjection.

Tingkat akurasi tanpa menggunakan fitur punctuation dan interjection yang didapatkan ialah sebesar 93%. Namun, jika data tersebut di tambahkan fitur punctuation dan interjection yang beberapa kali dilakukan, didapatkan tingkat akurasi sebesar 94% dengan jumlah interjection 100 kata, 93% dengan jumlah interjection 150 kata, dan 94% dengan jumlah interjection 200 kata.

Berdasarkan percobaan tersebut, tingkat akurasi terbaik yang didapatkan sebesar 94% melalui klasifikasi dengan jumlah interjection sebanyak 200 kata, walaupun memiliki tingkat akurasi yang sama dengan klasifikasi interjection dengan jumlah 100 kata, namun nilai dari precision maupun recall yang di tujukkan memberikan angka yang sedikit berbeda, yang mana hasil klasifikasi dengan interjection 200 kata memiliki nilai precision dan recall yang lebih besar.

5.Saran

Berdasarkan penelitian yang telah dilakukan oleh penulis telah di ambil kesimpulan seperti yang dijelaskan di atas, seperti halnya penelitian lainnya penelitian yang telah dilakukan masih memiliki beberapa kekurangan, sehingga penulis mengharapkan penelitian ini untuk dikembangkan lebih lanjut, saran yang dapat diberikan penulis berdasarkan hasil penelitian yaitu:

1. Banyaknya jumlah *dataset* sangat dalam penelitian ini, semakin banyaknya data yang digunakan justru akan semakin baik di dalam penelitian.
2. Pengacakan data adalah hal krusial yang dapat menyebabkan data terdistribusi secara natural.
3. Pembagian nilai label di dalam *dataset* yang kemudian di pecah kedalam bentuk data latih dan data uji lebih baik seimbang.
4. Algoritma klasifikasi yang di gunakan perlu mendapatkan algoritma klasifikasi pembandingan, terutama algoritma klasifikasi *deep learning*.

Referensi

- [1] R. WP, A. Novianty, and C. Setianingsih, "Analisis Sentimen Menggunakan Support Vector Machine dan Maximum Entropy," vol. 4, no. 2, p. 2389, 2017.
- [2] G. A. Buntoro, "Analisis Sentimen Hatespeech Pada Twitter Dengan Metode Naïve Bayes Classifier Dan Support Vector Machine," *J. Din. Inform.*, vol. 5, no. 2, pp. 1–12, 2016.
- [3] P. P. A. Ferdi Alvianda, Indriati, "Analisis Sentimen Konten Radikal Di Media Sosial Twitter Menggunakan Metode Support Vector Machine (SVM)," vol. 3, no. 1, pp. 7828–7833, 2019.
- [4] D. A. P. Rahayu, S. Kuntur, and N. Hayatin, "Sarcasm detection on Indonesian twitter feeds," *Int. Conf. Electr. Eng. Comput. Sci. Informatics*, vol. 2018-October, pp. 137–141, 2018.
- [5] B. Andrianto and S. Adinugroho, "Analisis Sentimen Konten Radikal Melalui Dokumen Twitter Menggunakan Metode Backpropagation," *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 2, no. 12, pp. 7380–7385, 2018.
- [6] A. Z. H. Khan, M. Atique, and V. . Thakare, "Sentiment analysis using Support Vector Machine," *I4CT 2014 - 1st Int. Conf. Comput. Commun. Control Technol. Proc.*, vol. 5, no. 4, pp. 333–337, 2014.
- [7] V. I. Santoso, G. Virginia, and Y. Lukito, "Penerapan Sentiment Analysis Pada Hasil Evaluasi Dosen Dengan Metode Support Vector Machine," *J. Transform.*, vol. 14, no. 2, p. 72, 2017.
- [8] O. Somantri, D. Apriliani, J. T. Informatika, P. Harapan, and B. Tegal, "Support Vector Machine Berbasis Feature Selection Untuk Sentiment Analysis Kepuasan Pelanggan Terhadap Pelayanan Support Vector Machine Based on Feature Selection for Sentiment Analysis Customer Satisfaction on Culinary," vol. 5, no. 5, pp. 537–548, 2018.
- [9] A. Perdana and M. T. Furqon, "Penerapan Algoritma Support Vector Machine (SVM) Pada Pengklasifikasian Penyakit Kejiwaan Skizofrenia (Studi Kasus: RSJ . Radjiman Wediodiningrat , Lawang)," *J. Pengemb. Teknol. Inf. dan Ilmu Komput. Univ. Brawijaya*, vol. 2, no. 9, pp. 3162–3167, 2018.
- [10] I. I Made Budi Surya Darma, Rizal Setya Perdana, "Penerapan Sentimen Analisis Acara Televisi Pada Twitter Menggunakan Support Vector Machine dan Algoritma Genetika sebagai Metode Seleksi Fitur," *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 2, no. 3, pp. 998–1007, 2018.

