

Analisis Sentimen Berbasis Leksikon InSet Terhadap Partai Politik Peserta Pemilu 2019 Pada Media Sosial Twitter

Ibnu Fanhar Nur. F.¹, Anisa Herdiani, S.T., M.T.², Widi Astuti, S.T., M.Kom.³

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

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

³widiwdu@telkomuniversity.ac.id

Abstrak

Pemilu adalah acara politik terbesar di dalam suatu bangsa, setiap partai akan berlomba-lomba memperebutkan suara rakyat. Untuk memperebutkan suara tersebut agar sesuai dengan target suatu partai maka dilakukanlah kampanye untuk menarik minat masyarakat terhadap suatu partai tertentu. Pada saat kampanye ini juga opini masyarakat sangat berkembang dengan masif terutama pada media sosial twitter yang mana masyarakat dapat mengekspresikan opininya secara bebas. Ada opini positif dan opini negatif serta opini yang bersifat netral yang dikeluarkan oleh masyarakat. Maka dari itu tugas akhir ini dibuat untuk menganalisis opini masyarakat terhadap partai politik peserta pemilu pada tahun 2019 dengan menggunakan InSet *Lexicon* dimana metode ini adalah metode berbasis kamus data berisi kata kata sentimen berbahasa Indonesia yang telah disertai oleh bobot pada setiap katanya. Berdasarkan sistem yang telah dibangun didapatkanlah rata rata pengujian sistem yang terdiri atas *precision 40%, Recall 42%, F1 35%, Accuracy 61%*.

Kata kunci : Pemilu, Partai Politik, Analisis Sentimen, Twitter, Leksikon.

Abstract

Elections are the biggest political event in a nation, each party will compete for the people's voices. To fight for these voices to fit the target of a party, a campaign was carried out to attract public interest in a particular party. During this campaign public opinion also developed massively, especially on Twitter social media, where the public could freely express their opinions. There are positive opinions and negative opinions as well as neutral opinions issued by the public. Therefore this final project was made to analyze public opinion of the political parties participating in the elections in 2019 using the *Lexicon InSet* where this method is a dictionary-based method of data containing Indonesian sentiment words which has been accompanied by a weight on each word. Based on the system that has been built, it is found that the average test system consists of *precision 40%, Recall 42%, F1 35%, Accuracy 61%*.

Keywords : Elections, Political Parties, Sentiment Analysis, Twitter, Lexicon

1. Pendahuluan

Latar Belakang

Secara umum pemilihan umum lahir dari konsep atau gagasan besar demokrasi, keadilan dan kesetaraan bagi seluruh individu disuatu negara dalam segala aspek kehidupan. Dalam pelaksanaan pemilu diperlukan peran aktif warga masyarakat untuk menyukseskan pemilihan umum ini. Masyarakat diberikan akses dan ruang untuk berperan aktif menjadi bagian dari sebuah pemilihan umum. Meskipun terkadang peran masyarakat masih cenderung hanya mengikuti prosedur dan momentum saja [1]. Maka atas hal itu dilakukanlah pemilu secara jujur, adil, langsung, umum, bebas dan rahasia setiap 5 (lima) tahun sekali [2].

Jika dilihat melalui data laporan yang diterbitkan oleh asosiasi penyelenggara jasa internet Indonesia (APJII) jumlah pengguna internet di Indonesia sebesar 171.176.716,8 orang dan 1.7% atau sekitar 291.000.417 orang adalah pengguna aktif dari media sosial twitter [3]. Selain itu media sosial juga telah dijadikan sebuah strategi komunikasi politik yaitu digunakan untuk menyebarkan visi dan misi suatu partai yang akan berpartisipasi di dalam sebuah pemilu. Kampanye di media sosial juga dinilai sebagai sarana komunikasi yang tepat karena akan berperan membawa orang (penggunanya) untuk berpartisipasi secara aktif dengan memberi kontribusi dan *feedback* secara terbuka, baik untuk membagi informasi maupun memberi respon secara online dalam waktu yang cepat [4]. Banyaknya kontribusi atau *feedback* yang berupa opini masyarakat terhadap partai politik di media sosial ini belum mendapatkan pemastian atau pengujian [5].

Dari permasalahan di atas inilah dibutuhkan sebuah sistem yang dapat melakukan analisis sentimen yang berkembang di media sosial terutama twitter. Pada penelitian ini menggunakan pendekatan *lexicon resource* berbahasa Indonesia sehingga memiliki keunggulan tidak perlu melakukan translasi kedalam Bahasa Inggris. Sentimen leksikon berbahasa Indonesia yang digunakan pada penelitian ini pun telah memiliki bobot nilai atau *polarity score* pada setiap katanya. Dimana untuk mendapatkan *polarity score* harus melalui tahapan matematis terlebih dahulu sehingga cukup baik untuk digunakan dalam melakukan analisis sentimen. *Polarity score* ini digunakan untuk mengklasifikasikan jenis sentimen sehingga pada hasil akhir akan didapatkan apakah kalimat tersebut bersentimen positif, negatif atau netral [6]

Topik dan Batasannya

Tugas akhir ini memiliki topik tentang analisis sentimen atau biasa disebut *opinion mining*. Analisis sentimen ini biasanya dilakukan dengan cara *crawling* data melalui media sosial twitter. Karena media sosial twitter ini sering sekali digunakan dalam sarana kampanye dan di dalam twitter juga hanya mampu menampung 140 karakter yang berarti kalimat yang muncul dapat langsung diketahui maksud dan tujuannya selain itu juga masyarakat Indonesia adalah pengguna twitter aktif paling banyak pada tahun 2015 [7]. Maka dari itu analisis sentimen ini menggunakan sarana media sosial twitter untuk mengambil data awal penelitian. Perumusan masalah dalam penelitian ini sebagai berikut :

1. Bagaimana mengklasifikasikan jenis sentimen dari suatu kalimat *tweet* dengan menggunakan InSet *lexicon* berbahasa Indonesia?
2. Bagaimana tingkat performansi dari metode leksikon dalam melakukan klasifikasi sentiment?

Selain rumusan masalah, penelitian ini juga memiliki beberapa keterbatasan dalam pelaksanaan. Batasan masalah penelitian ini diantaranya :

1. Dataset hanya berdasarkan *tweet* pada media sosial *twitter* yang telah dilakukan *crawling*.
2. *Tweet* yang dicrawling seputar pemilu 2019.
3. *Tweet* partai yang diambil adalah partai yang memiliki legalitas secara nasional (bukan daerah tertentu saja).
4. Metode pengambilan dataset adalah menggunakan @namapartai, #namapartai, dan nama partai secara langsung
5. Pencocokan kata dalam proses penghitungan bobot hanya melihat kata per kata.

Tujuan

Selain menentukan rumusan masalah dan Batasan masalah penelitian ini juga menentukan tujuan, agar hasil penelitian sesuai dengan apa yang sudah direncanakan. Berikut adalah beberapa tujuan dari penelitian ini :

1. Dapat menghitung bobot atau *polarity score* dari suatu kalimat dan menentukannya ke dalam jenis sentimen.
2. Dapat memberikan perhitungan performansi sistem dengan menggunakan *Precision, Recall, F1 dan accuracy* untuk hasil analisis sentiment dengan menggunakan InSet *lexicon* berbahasa Indonesia.

2. Studi Terkait

Lexicon Based Approach

Lexicon based adalah metode ilmiah yang sering digunakan dalam suatu penelitian analisis sentimen. Cara kerja metode ini adalah dengan menggunakan sebuah kamus kata atau korpus yang dilengkapi dengan bobot pada setiap katanya sebagai sumber Bahasa atau leksikal. Hasil analisis dengan metode ini adalah berupa klasifikasi sentimen positif, negatif dan netral [8]. Metode ini adalah bagian dari *machine learning* yang bersifat *unsupervised*. Kualitas dari hasil tergantung pada kamus kata atau *corpus* yang digunakan. Penggunaan metode ini cukup mudah, yaitu bobot dalam kata yang berhasil ditemukan oleh sistem akan dihitung dan hasilnya akan diklasifikasikan kedalam sebuah sentimen. Dalam melakukan klasifikasi, metode ini memiliki aturan umum yaitu,

- | | | |
|---------------------------------------|------------------|-----|
| <i>If sentiment score > 0 then</i> | Sentimen Positif | (1) |
| <i>If sentiment score < 0 then</i> | Sentimen Netral | (2) |
| <i>If sentiment score = 0 then</i> | Netral. | (3) |

Sebagai contoh : “Saya ingin[2] bisa[2]memiliki keahlian[4] khusus”. Pada kalimat contoh tersebut kata yang berhasil ditemukan adalah ingin, bisa dan khusus. Lalu selanjutnya bobot dari kata tersebut dijumlahkan (2+2+4 = 8) maka kalimat tersebut memiliki sentimen positif [9].

Preprocessing

Text processing atau biasa disebut *preprocessing* adalah tahap awal dari sebuah analisis berbasis *text*. Hal ini dilakukan untuk mempersiapkan *text* menjadi data yang akan siap diolah lebih lanjut. [10].

Analisis Sentimen

Analisis sentimen atau *opinion mining* merupakan proses memahami, mengestrak dan mengolah data secara otomatis untuk mendapatkan sebuah informasi sentimen atau opini yang terkandung dalam suatu kalimat. Analisis sentimen dilakukan untuk melihat pendapat atau kecenderungan opini terhadap suatu masalah ataupun objek oleh seseorang. *Bentuk dari opini bisa berupa opini yang positif, negatif atau netral* [11]. Secara umum analisis sentimen dibagi menjadi 3 (tiga) jenis yaitu [12] :

- a. Analisis sentimen pada level kalimat
Tujuan analisis sentimen pada level kalimat ini bertujuan untuk mengklasifikasikan kalimat kedalam jenis positif, negatif dan netral.
- b. Analisis sentimen pada level dokumen
Tujuan analisis sentimen pada level dokumen ini adalah mengklasifikasikan dokumen secara keseluruhan.
- c. Analisis sentimen berbasis fitur
Merupakan analisis paling detail terhadap suatu fitur yang terdapat pada suatu objek.

Crawling

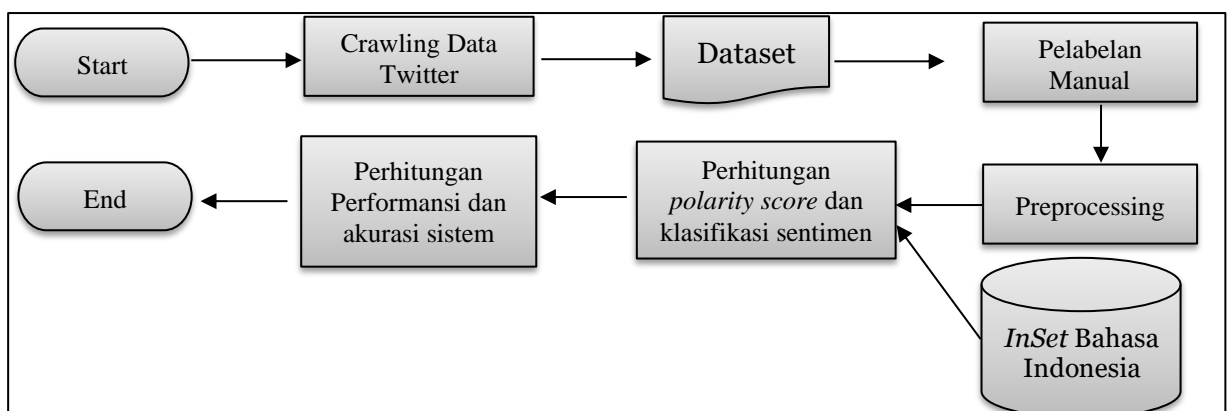
Crawling adalah suatu teknik yang digunakan untuk mengumpulkan data yang terdapat di dalam suatu *web*. *Crawling* bekerja secara otomatis, dimana informasi yang akan didapatkan atau dikumpulkan adalah sesuai dengan kata kunci yang telah diberikan oleh pengguna. Media yang digunakan untuk melakukan *crawling* ini adalah *crawler*. Dimana *crawler* ini berbentuk program yang dibuat dengan algoritma tertentu sehingga dapat melakukan pemindaian terhadap website yang ingin di *crawl* dengan kata kunci yang telah diberikan sebelumnya. Hasil dari *crawling* tersebut akan tersimpan di dalam file atau kedalam database yang telah disiapkan sebelumnya [13]. Banyak kata kunci yang dapat digunakan untuk melakukan *crawling* data dari media sosial twitter ini. Namun kata kunci atau *keyword query* yang digunakan tidak lebih dari 10 kata, beberapa kata kunci yang umum digunakan seperti dengan menyebutkan langsung objek yang dicari, selanjutnya dengan mencari nama akun, dan menggunakan tagar atau *hashtag*. Contoh kata kunci yang dapat digunakan untuk *crawling* dapat dilihat pada tabel 1 [14].

Tabel 1 : Contoh Kata Kunci Untuk *Crawling*

Jenis Kata Kunci	Contoh Masukan	Keterangan
Menyebut langsung objek	partai	Untuk mencari <i>tweet</i> yang mengandung kata “partai”
Mencari nama akun	@partai	Untuk mencari <i>tweet</i> yang mengandung “@partai”
Dengan tagar atau <i>hashtag</i>	#partai	Untuk mencari <i>tweet</i> yang mengandung <i>hashtag</i> partai

3. Sistem yang Dibangun

Perancangan sistem yang dibangun adalah berdasarkan hasil studi literatur dan berdasarkan kebutuhan serta desain sistem yang telah terdefinisi sebelumnya. Sistem yang akan dibangun yaitu sebuah sistem yang dapat melakukan Analisa sentimen terhadap komentar atau opini masyarakat terhadap partai partai peserta pemilu 2019. Untuk lebih jelasnya dapat dilihat pada gambar 1 :



Gambar 1 : Alur Kerja Sistem

3.1 Crawling Data Twitter

Sistem yang dibangun diawali dengan melakukan *crawling* data. *Crawling* berproses mencari data sesuai dengan masukan pengguna yang akan dijadikan dataset untuk keperluan penelitian ini. *Keyword crawling* data yang telah dilakukan untuk penelitian ini dapat dilihat pada tabel 2.

Tabel 2 : Masukan Untuk Proses *Crawling* Data

No	Nama Partai	Singkatan Partai	Menggunakan @	Menggunakan #	Menggunakan Singkatan Partai
1	Partai Kebangkitan Bangsa	PKB	@DPP_PKB	#PKB	PKB
2	Partai Gerindra	Gerindra	@Gerindra	#Gerindra	Gerindra
3	Partai Demokrasi Indonesia Perjuangan	PDIP	@PDI_Perjuangan	#PDIPerjuangan	PDIP
4	Partai Golongan Karya	Golkar	@PartaiGolkar	#Golkar	Golkar
5	Partai Nasional Demokrat	Nasdem	@NasDem	#Nasdem	Nasdem
6	Partai Garuda	Garuda	@DPPPartaiGaruda	#PartaiGaruda	Garuda
7	Partai Berkarya	Berkarya	@BerkaryaOfc	#Partaiberkarya	Berkarya
8	Partai Keadilan Sejahtera	PKS	@PKSejahtera	#PKS	PKS
9	Partai Persatuan Indonesia	Perindo	@PartaiPerindo	#Perindo	Perindo
10	Partai Persatuan Pembangunan	PPP	@DPP_PPP	#PPP	PPP
11	Partai Solidaritas Indonesia	PSI	@psi_id	#PartaiPSI	PSI
12	Partai Amanat Nasional	PAN	@Official_PAN	#PAN	PAN
13	Partai Hati Nurani Rakyat	Hanura	@PartaiHANURA	#Hanura	Hanura
14	Partai Demokrat	Demokrat	@PDemokrat	#Demokrat	Demokrat
15	Partai Keadilan Dan Persatuan Indonesia	PKPI	@pkp_id	#PKPI	PKPI
16	Partai Bulan Bintang	PBB	@PBB2019	#PBB	PBB

3.2 Dataset

Dataset dari penelitian ini bersumber dari hasil *crawling* melalui media sosial twitter pada rentang April 2019. Topik yang dicari adalah mengenai pemilu 2019. Proses pencarian dataset telah disebutkan pada poin sebelumnya. Dataset yang terkumpul pada penelitian ini sejumlah 1724 baris yang berasal dari cuitan masyarakat mengenai pemilu 2019. Contoh dataset pada penelitian ini dapat dilihat pada tabel 3 berikut ini :

Tabel 3 : Contoh Dataset

b'@sfi4ns Jman pak harto yg mempersekusi perangkat negara tpi orang2 golkar tdk mlakukannya tpi skarang selain perangkat negara juga para pendukung/pnggembira ikut nlakukannya kan edan'
b'@yervelveu Nyanyiin mars golkar dulu nder'

3.3 Pelabelan Manual

Pada proses pelabelan data ini dilakukan oleh 7 orang, dimana sistem pelabelan menggunakan metode kualitatif. Yaitu sebuah metode survei yang mengharuskan seluruh pertanyaan dari kuisisioner tersebut terisi dengan baik dan sesuai dengan instruksi. Pada pelabelan ini para pelabel akan melakukan Analisa kalimat per kalimat dan memberikan penilaian jenis sentimen apakah yang terkandung di dalamnya. Jenis penilaian sentimen pada penelitian ini berupa positif yang ditandai dengan angka 1, negatif yang ditandai dengan angka 0, dan netral yang ditandai dengan huruf x. Komposisi dari pelabel manual ini terdiri atas :

- Mahasiswa *Digital Public Relation* sebanyak 2 orang
- Mahasiswa Ilmu Komunikasi sebanyak 1 orang
- Mahasiswa Informatika sebanyak 2 orang
- Wartawan bidang politik sebanyak 1 orang
- Dan wiraswasta sebanyak 1 orang

Pelabel manual dalam hal ini memiliki keahlian dalam membedakan kalimat pada dataset untuk diteliti jenis sentimennya secara manual.

3.4 Preprocessing

Proses ini dilakukan sebelum penelitian membangun ontologi, dimana *preprocessing* dilakukan untuk mempermudah melakukan pengolahan data. Tahap tahap *preprocessing* yang dilakukan pada penelitian ini adalah sebagai berikut :

1. Menghapus Simbol

Menghapus Simbol adalah proses dimana pada kalimat yang akan diolah dilakukan penghapusan simbol simbol yang tidak memiliki makna penting .

Contoh proses menghapus simbol dapat dilihat pada tabel 4.

Tabel 4 : Contoh Proses Menghapus Simbol

Input	Output
b'@. berkaryaofc pemimpin yang visioner #satgasmandala'	BerkaryaOFCpemimpin yang visioner

2. Casefold

Casefold adalah proses dimana seluruh huruf pada kalimat yang akan diolah diubah menjadi huruf kecil pada keseluruhan kalimatnya. Contoh proses *casefold* dapat dilihat pada tabel 5.

Tabel 5 : Contoh Proses *Casefold*

Input	Output
BerkaryaOFC pemimpin yang visioner	berkaryaofc pemimpin yang visioner

3. Tokenisasi

adalah proses dimana seluruh kata pada kalimat yang akan diolah dipisahkan kata perkata. Contoh proses Tokenisasi dapat dilihat pada tabel 6.

Tabel 6 : Contoh Proses Tokenisasi

Input	Output
berkaryaofc pemimpin yang visioner	'berkaryaofc' , 'pemimpin' , ' yang ' , 'visioner'

3.5 Perhitungan *Polarity Score* Dan Klasifikasi Sentimen

Perhitungan *polarity score* dalam analisis ini didasarkan oleh *InSet lexicon* yang telah didapatkan sebelumnya. *InSet lexicon* ini berisi kata kata positif dan negatif beserta bobot dari kata tersebut. Contoh *InSet lexicon* positif dapat dilihat pada tabel 7.

Tabel 7 : *InSet Lexicon* Positif

Kata	Bobot
Adil	4
Jujur	4
Lelap	4
Tembus	3
Lipur	2
Sajak	2
Reaksi	2

Dan untuk *InSet lexicon* negatif dapat dilihat pada tabel 8.

Tabel 8 : *InSet Lexicon* Negatif

Kata	Bobot
Gila	-4
Sinting	-5
Aduh	-3
Fiksi	-2
Terlalu	-3
Gendeng	-2
Duh	-4

Kamus tersebut digunakan untuk mencocokkan kata beserta bobot yang terkandung untuk selanjutnya dilakukan perhitungan pada setiap kalimat. Proses perhitungan *polarity score* dilakukan dengan cara menjumlahkan keseluruhan bobot dari kata yang terdeteksi oleh sistem dan selanjutnya akan diklasifikasikan kedalam jenis sentimen. Secara umum dinyatakan dengan kondisi sebagai berikut :

If sentiment score > 0 then Sentimen Positif (4)

If sentiment score < 0 then Sentimen Netral (5)

If sentiment score = 0 then Netral (6)

Pada penelitian ini juga, pengkondisian *If sentiment score = 0 then* Netral adalah kalimat yang memiliki kata dan bobot yang terdeteksi namun memiliki hasil akhir *sentiment score* 0. Pada proses ini juga, jika suatu kata terdapat pada kelas positif dan negatif maka kata tersebut tidak dicantumkan dalam perhitungan *polarity score*. Contoh perhitungan *polarity score* dapat dilihat pada tabel 9.

Tabel 9 : Perhitungan *Polarity Score*

Tweet	Kata	Score	Total Score
ya ampun brisik banget nih kamu	/ ya	/4	4
kader partai odong odong ngak usah atur partai besar pikirin aja gimana partai odong odong mu bisa populer malu malu aja kau jadi kader	/ aja / populer / usah	/1/5/-5	1
aah cangcut merah mah sdh lepas kendali dr itu mnunjukkan ketidk brdayaan partai bs jg brskap abu 2	/ itu / lepas	/-2/-2	-4
tapi demokrat suaranya nyungsep			0
hidup 01 yess			0
hidup 01			0
klo pan sma pks masih ada kemungkinan kecil berani tp klo demokrat g yakin paling pilih main 2 kaki wkwwk	/ wkwwk / yakin / kemungkinan / pilih / kecil	/3/3/-1/-2/-3	0
pemilu serentak baru kali ini boss n nlgian emang ada yg melihat dan mengkoreksi kaya sekarang	/ serentak / melihat	/3/3	6
lebih baik nya di evaluasi seluruhnyaaa	/ lebih	/1	1
mantap	/ mantap	/5	5
perlu jiwa ksatria untuk mengakui sebuah kekalahan n n tco	/ perlu	/2	2

3.6 Perhitungan Performansi Dan Akurasi Sistem

Pada tahap ini digunakan teknik *precision*, *recall*, F1 dan *accuracy* untuk menghitung nilai performansi dari sistem. Perhitungan ini membutuhkan empat komponen yaitu TP (*true positive*), TN (*true negative*), FP (*false positive*), FN (*false negative*). Perhitungan performansi ini dilakukan pada setiap kelas jenis sentiment terlebih dahulu kemudian dilakukan perhitungan rata rata. Jika digambarkan dalam tabel maka empat komponen tersebut adalah seperti pada tabel 10 berikut :

Tabel 10 : *Confusion Matrix*

	Predicted class (YES)	Predicted class (NO)
Actual Class (YES)	TP	FN
Actual Class (NO)	FP	TN

Sedangkan rumus dari masing masing perhitungan dapat dilihat pada persamaan berikut :

$$Precision = \frac{TP}{TP + FP} \quad (7)$$

$$Recall = \frac{TP}{TP + FN} \quad (9)$$

$$F1 = \frac{2(\text{precision} \times \text{recall})}{\text{precision} + \text{recall}} \quad (8)$$

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \quad (10)$$

Keterangan :

1. *Precision* merupakan perhitungan jumlah kemunculan data yang diklasifikasikan true dengan benar dibanding jumlah data yang diklasifikasikan true.
2. *Recall* merupakan jumlah kelas yang diklasifikasikan dengan benar pada sebuah kelas dibanding seluruh data hasil klasifikasi sistem pada kelas tersebut
3. F1 / *F-Measure* merupakan hasil rata rata antara *precision* dan *recall*
4. *Accuracy* merupakan rasio prediksi benar dengan keseluruhan data.

4. Evaluasi

4.1 Hasil Pengujian

Setelah melalui tahapan pembangunan sistem, dilakukan pengujian terhadap sistem. Pengujian ini bertujuan untuk mengetahui seberapa baik kinerja sistem yang dibangun untuk mengklasifikasikan sentimen terhadap seluruh tweet di dalam dataset. Contoh hasil yang didapat dari pengujian sistem dapat dilihat pada tabel 11.

Tabel 11 : Hasil Pengujian Sistem

Tweet	Kata	Score	Total Score	Prediksi	Aktual
ya ampun brisik banget nih kamu	/ ya	/4	4	POSITIF	NETRAL
kader partai odong odong ngak usah atur partai besar pikirin aja gimana partai odong odong mu bisa populer malu maluin aja kau jadi kader	/ aja / populer / usah	/1/5/-5	1	POSITIF	NEGATIF
aah cangcut merah mah sdh lepas kendali dr itu mnunjukkan ketidk brdayaan partai bs jg brskap abu 2	/ itu / lepas	/-2/-2	-4	NEGATIF	NEGATIF
tapi demokrat suaranya nyungsep	-	-	-	-	NEGATIF
hidup 01 yess	-	-	-	-	POSITIF
hidup 01	-	-	-	-	POSITIF

Tweet	Kata	Score	Total Score	Prediksi	Aktual
lebih baik nya di evaluasi seluruhnyaaa	/ lebih	/1	1	POSITIF	POSITIF
mantap	/ mantap	/5	5	POSITIF	POSITIF
perlu jiwa ksatria untuk mengakui sebuah kekalahan n n tco	/ perlu	/2	2	POSITIF	POSITIF
orang stres partai gak masuk ngoceh aja	/ aja / stres / masuk	/1/-4/-3	-6	NEGATIF	NEGATIF
mengatakan siap bekerjasama dengan siapapun termasuk pemerintah terpilih nantinya untuk mensukseskan 14 program prioritas demokratan n	/ siap / mengatakana / termasuk	/1/3/-3	1	POSITIF	POSITIF
klo pan sma pks masih ada kemungkinan kecil berani tp klo demokrateg yakin paling pilih main 2 kaki wkwkwk	/ wkwkwk / yakin / kemungkinan / pilih / kecil	/3/3/-1/-2/-3	0	NETRAL	NEGATIF
pemilu serentak baru kali ini boss n nlgian emang ada yg melihat dan mengkoreksi kaya sekarang	/ serentak / melihat	/3/3	6	POSITIF	NETRAL
setelah 22 mei sby dan demokratt masuk koalisi jokowidrama pilpres selesai	/ masuk / drama	/-3/-4	-7	NEGATIF	NEGATIF

Dari total data sejumlah 1.723, data yang berhasil dilakukan prediksi oleh sistem adalah sebesar 1.493. sedangkan sejumlah 230 tidak dapat diprediksi oleh sistem.

Setelah dilakukan pengujian terhadap sistem, yaitu proses uji performansi terhadap sistem yang telah dilakukan pengujian. Uji pada penelitian ini dilakukan untuk melihat performansi dari metode *lexicon* yang dilakukan untuk melakukan analisis sentimen masyarakat terhadap partai politik peserta pemilu 2019 di twitter. Sehingga terdapat hasil pengujian yaitu *precision*, *recall*, F1 dan *accuracy* pada tabel 12.

Tabel 12 : Hasil Performansi Sistem

Jenis Sentimen	Precison	Recall	F1	Accuracy
Positif	44%	42%	43%	66%
Negatif	40%	79%	53%	53%
Netral	37%	5%	9%	63%

Setelah didapatkan hasil Performansi sistem berdasarkan jenis sentimen, maka selanjutnya adalah merata-rata hasil pengujian sistem tersebut. Rata-rata hasil pengujian sistem dapat dilihat di tabel 13.

Tabel 13 : Hasil Performansi Sistem

Nilai Rata Rata Performansi			
Precision	Recall	F1	Accuracy
40%	42%	35%	61%

4.2 Analisis Hasil Pengujian

Melihat hasil pengujian dan performasi dari sistem maka dapat dilihat bahwa :

1. Masih terdapat data yang belum bisa dilakukan prediksi oleh sistem, hal ini karena beberapa faktor, diantaranya :
 - a. Kata yang terkandung di dalam dataset berupa singkatan sehingga sistem tidak dapat menemukan kecocokan dari dataset terhadap data kamus kata. Contoh bahwa sistem tidak dapat menemukan kata dalam InSet dapat dilihat pada tabel 14.

Tabel 14 : Contoh Hasil Sistem Tidak Dapat Menemukan Kata Singkatan

Tweet	Kata	Score	Total Score	Prediksi	Aktual
tdk	-	-	-	-	NETRAL

- b. Kata yang terkandung di dalam dataset tidak ada di dalam kamus kata. Contoh bahwa sistem tidak dapat menemukan kata karena tidak dalam kamus kata dapat dilihat pada tabel 15.

Tabel 15 : Contoh Hasil Sistem Tidak Dapat Menemukan Kata Karena Tidak Dalam Kamus Kata

Tweet	Kata	Score	Total Score	Prediksi	Aktual
dengan tanpa demokrat saluud jokowi	-	-	-	-	POSITIF
udah biasa demokrat mlm banyak kaki	-	-	-	-	NEGATIF
partai demokrat hahahaha hahahaha hahheeyyyyy	-	-	-	-	NETRAL

- c. Kata yang terkandung di dalam dataset berada di dua kelas (positif dan negatif) sehingga dilakukan penghapusan agar tidak menjadi keambiguan dan menyebabkan kesalahan hasil prediksi sistem. Contoh kata yang berada di dua kelas dapat dilihat pada tabel 16. Untuk contoh hasil sistem yang menghapus kata di dua kelas terdapat pada tabel 17 dan untuk contoh sistem yang tidak menghapus kata di dua kelas terdapat pada tabel 18.

Tabel 16 : Contoh Kata Yang Berada Di Dua Kelas

Kata Positif	Bobot	Kata Negatif	Bobot
Doang	1	Doang	-1
Adil	4	Adil	-3
Suka	3	Suka	-1

Tabel 17 : Contoh Hasil Sistem Yang Menghapus Kata Di Dua Kelas

Tweet	Kata	Prediksi	Aktual
dan jujur aja suka gagasan sim seumur hidup pajak stnk gratis nsemoga bukan wacana doang ya	/ gratis / aja / ya / wacana / pajak / bukan	NETRAL	NETRAL

Tabel 17 : Contoh Hasil Sistem Yang Tidak Menghapus Kata Di Dua Kelas

Tweet	Kata	Prediksi	Aktual
dan jujur aja suka gagasan sim seumur hidup pajak stnk gratis nsemoga bukan wacana doang ya	/ doang / wacana / jujur / pajak / suka / bukan / hidup / gratis / doang / jujur / aja / suka / hidup / ya	POSITIF	NETRAL

Berdasarkan contoh diatas, terdapat kesalahan prediksi terhadap *tweet* yang tidak dilakukan penghapusan deteksi untuk kata yang berada di dua kelas. Hal ini tentu sangat berdampak pada hasil prediksi sistem

2. Hasil Pengujian Pada kelas positif

Precision : Sistem hanya mampu menemukan prediksi yang tepat sejumlah 190 *tweet* yang sesuai dengan data aktual, sedangkan sistem melakukan prediksi yang tidak sesuai data aktual sejumlah 238 *tweet*.

Recall : Sistem mampu memprediksi sejumlah 190 *tweet* yang relevan dengan data aktual, sedangkan sistem memprediksi 263 *tweet* aktual yang seharusnya positif menjadi bukan positif.

Dari 2 pengujian sistem tersebut dilakukan rata-rata sehingga mendapatkan nilai F1 sebesar 43% dan menghasilkan akurasi sebesar 66%. Yang berarti bahwa sistem dapat melakukan prediksi yang tepat sesuai data aktual lebih dari 50% data.

3. Hasil Pengujian Pada kelas negatif

Precision : Sistem hanya mampu menemukan prediksi yang tepat sejumlah 399 *tweet* yang sesuai dengan data aktual, sedangkan sistem melakukan prediksi yang tidak sesuai data aktual sejumlah 591 *tweet*.

Recall : Sistem mampu memprediksi sejumlah 399 *tweet* yang relevan dengan data aktual, sedangkan sistem memprediksi 107 *tweet* aktual yang seharusnya negatif menjadi bukan negatif.

Dari 2 pengujian sistem tersebut dilakukan rata-rata sehingga mendapatkan nilai F1 sebesar 53% dan menghasilkan akurasi sebesar 53%. Yang berarti bahwa sistem dapat melakukan prediksi yang tepat sesuai data aktual lebih dari 50% data.

4. Hasil Pengujian Pada kelas netral

Precision : Sistem hanya mampu menemukan prediksi yang tepat sejumlah 28 *tweet* yang sesuai dengan data aktual, sedangkan sistem melakukan prediksi yang tidak sesuai data aktual sejumlah 47 *tweet*.

Recall : Sistem mampu memprediksi sejumlah 28 *tweet* yang relevan dengan data aktual, sedangkan sistem memprediksi 506 *tweet* aktual yang seharusnya netral menjadi bukan netral.

Dari 2 pengujian sistem tersebut dilakukan rata-rata sehingga mendapatkan nilai F1 sebesar 9% dan menghasilkan akurasi sebesar 63%. Yang berarti bahwa sistem dapat melakukan prediksi yang tepat sesuai data aktual lebih dari 50% data.

Dari ketiga hasil pengujian performansi sistem dilakukan perhitungan rata-rata untuk mendapatkan rata-rata performansi keseluruhan sistem. Dan didapatkan hasil rata-rata *precision* 40%, *Recall* 42%, *F1* 35% dan *Accuracy* 61%.

5. Kesimpulan

Setelah melalui proses *crawling* dengan menggunakan 3 (tiga) jenis kata kunci lalu melakukan pelabelan manual hingga melakukan perhitungan performansi akurasi sistem. Dapat disimpulkan bahwa, data yang digunakan untuk analisis ini sudah banyak. Pada analisis ini, data yang digunakan sebagai dataset sejumlah 1.723 *tweet*. Setelah seluruh dataset dilakukan analisis dengan mencocokkan kata pada korpus inset yang ada maka didapatkan bahwa data yang berhasil memiliki prediksi sebanyak 1.493 yang berarti sebanyak 86% data berhasil diproses oleh sistem. Hasil proses sistem ini kemudian dilakukan perhitungan performansi akurasinya. Dan hasilnya adalah Dan di dapatkan hasil rata-rata *precision* 40%, *Recall* 42%, *F1* 35% dan *Accuracy* 61%. Hasil ini cukup baik namun dengan beberapa catatan agar nantinya sistem dapat mendapatkan hasil *precision*, *recall*, *F1*, dan *accuracy* lebih baik lagi.

5.1 Saran

Hal – Hal yang perlu dipertimbangkan kedepannya adalah sebagai berikut :

- a. Melakukan *preprocessing* yang dapat melakukan *handling* terhadap sebuah singkatan kata. Sehingga dataset akan bebas dari kata kata yang disingkat dan hasil Analisa akan menjadi lebih baik.
- b. Memperkaya kata di dalam inset. Hal ini bertujuan agar semakin banyak kata yang berhasil dideteksi oleh sistem pada saat melakukan perhitungan *polarity score*. Karena semakin banyak kata yang terdapat dalam inset maka akan semakin bagus sebuah sistem dalam melakukan prediksi
- c. Membuat sebuah sistem yang dapat melakukan prediksi terhadap konteks suatu kalimat. Hal ini agar tidak perlu dilakukan penghapusan kata yang berada di dua kelas namun sistem dapat memilih kata mana yang cocok untuk digunakan sebagai prediksi suatu kalimat tersebut.

Daftar Pustaka

- [1] F. R. Bachtiar, "PEMILU INDONESIA: KIBLAT NEGARA DEMOKRASI DARI BERBAGAI REFRESENTASI," *Jurnal Politik Profetik*, vol. 3, no. 1, 19 December 2014.
- [2] P. R. Indonesia, "UNDANG-UNDANG REPUBLIK INDONESIA NOMOR 3 TAHUN 1999," Dewan Perwakilan Rakyat, Jakarta, 1999.
- [3] Polling Indonesia, "Laporan Survei Penetrasi & Profil Perilaku Pengguna Internet Indonesia," Asosiasi Penyelenggara Jasa Internet Indonesia, Jakarta, 2018.
- [4] B. I. M. Ike Atikah Ratnamulyani, "PERAN MEDIA SOSIAL DALAM PENINGKATAN PARTISIPASI PEMILIH PEMULA DIKALANGAN PELAJAR DI KABUPATEN BOGOR," *Jurnal Ilmu-ilmu Sosial dan Humaniora* , vol. 20, no. 2, p. 155, 2018.
- [5] F. Ratnawati, "Implementasi Algoritma Naive Bayes Terhadap Analisis Sentimen Opini Film Pada Twitter," *JURNAL INOVTEK POLBENG -SERI INFORMATIKA*, vol. 3, no. 1, p. 50, 2018.
- [6] M. LAILIYAH, "Sentiment Analysis Menggunakan Rule Based Method Pada Data Pengaduan Publik Berbasis Lexical Resources," in *TESIS - TE14259*, Institut Teknologi Sepuluh November, 2017, p. VII.
- [7] R. S. P. M. A. F. Agnes Rossi Trisna Lestari, "Analisis Sentimen Tentang Opini Pilkada Dki 2017 Pada Dokumen Twitter Berbahasa Indonesia Menggunakan Naive Bayes dan Pembobotan Emoji," *Journal Pengembangan Teknologi Informasi dan Ilmu Komputer*, vol. 1, no. 12, p. 1718, 2017.
- [8] C. Sutami, "PERBANDINGAN METODE KLASIFIKASI NAIVE BAYES CLASSIFIER DAN LEXICON BASED DALAM ANALISIS SENTIMEN (Studi Kasus : Twitter)," Fakultas Teknik Universitas Widyatama, Bandung, 2015.
- [9] T. T. P. S. P. C. T. T. Olga Kolchyna, "Methodology for Twitter Sentiment Analysis," *Twitter Sentiment Analysis: Lexicon Method, Machine Learning Method and Their Combination*, 2015.
- [10] R. Larasati, "KLASIFIKASI TEKS DENGAN MENGGUNAKAN ALGORITMA K-NEAREST NEIGHBOR PADA DOKUMEN TUGAS AKHIR," *Tugas Akhir*, no. Universitas Widyatama, 2015.
- [11] S. H. P. E. A. D. Imam Fahrur Rozi, "Implementasi Opinion Mining (Analisis Sentimen) untuk Ekstraksi Data Opini Publik pada Perguruan Tinggi," *Jurnal EECCIS* , vol. 6, p. 37, 2012.
- [12] I. B. Yislam, "Analisis Sentimen Masyarakat Terhadap Pemerintahan Jokowi Menggunakan Data Twitter," no. Universitas Indonesia.
- [13] P. Y. Saputra, "IMPLEMENTASI TEKNIK CRAWLING UNTUK PENGUMPULAN DATADARI MEDIA SOSIAL TWITTER," *Jurnal Dimanika Dotcom*, vol. 8, no. 2, p. 162, 2017.
- [14] E. B. S. Z. A. B. Jaka Eka Sembodo, "Data Crawling Otomatis pada Twitter," *INDOSC*, p. 11, 2016.