

BAB V

HASIL PENELITIAN DAN PEMBAHASAN

Ujicoba penelitian ini dilakukan dengan dataset yang terdiri dari tweet berupa opini berjumlah 2000 buah yang di bagi berdasar dua calon kandidat presiden. Data diambil secara manual dengan syntax python dan disimpan dalam file CSV. Pengambilan data dilakukan menggunakan query pencarian seperti pada tabel dibawah.

Hastagh	Jumlah
#jokowi	1000
#prabowo	1000

5.1. Proses Akuisisi Data

Table 3 AKUISISI DATA

NO	TWEET
1	Dikit dikit tagar dikit dikit tagar, tagar kok dikit dikit#KPUNgejarQuickCount #PenjarakanAndreTaulany #Jokowi... https://t.co/mMmcJtSAGQ
2	Sakdiyah Ma'ruf: Indonesia Penuh Lelucon Politik, Saatnya Komedi Memimpin #jokowi #berSATUikutKyai https://t.co/jZjRbfxXL9
3	Ketegangan dalam bekerja akan menimbulkan dampak yg besar bagi kesehatan kita mgkn itu yg terjadi pada para pejuang... https://t.co/NXzf71riqp
4	Untuk alasan apapun, ketua MPR harus diganti dan biarkan PAN jadi oposisi sampai 2024.#savekpu #kpu... https://t.co/rWACCS9z6p
5	One Belt One Road (OBOR), Modus Invansi Senyap Ala Cina Klik https://t.co/ci2bJF2ugV #obor #jalur #sutera #cina... https://t.co/LHtgml5ZMi
6	Soal Peluang Demokrat Gabung KIK, Ferdinand: Tergantung Pak Jokowi https://t.co/CYJDNR7yJW #pantaucom... https://t.co/ovdJO8nzLM
7	Prabowo 'Ngambek' karena AHY Penuhi Undangan Jokowi? Andi Arief Berkicau https://t.co/uiWNfL3BeY #Jokowi #AHY #Prabowo #Netarlnews
8	Siapaapun yang terpilih harus menjadi Kemenangan rakyat Indonesia. #2019pilihjokowi #jokowi #emakemakjamannow... https://t.co/8KV4YHxEBE

9	Ketua DPC Gerindra Apresiasi Kinerja KPU dan Bawaslu Bangkalan #jokowi #berSATUikutKyai https://t.co/C3fmFVh1Fj https://t.co/RSvpdcn8VI
10	Grace Natalie Tak #Pantas Jadi Menteri Kabinet #Jokowi https://t.co/HRiX4eY40j #moreshareforum #beritaterkini
11	selamat dan semoga Indonesia semakin Maju #jokowi #KPUSelaluTransparan https://t.co/5t4h3NfeYu
12	mantap Pak Jokowi dan Kyai Ma'ruf, selamat ya #jokowi #KPUSelaluTransparan https://t.co/wiUZPBNdqT
13	mantap bgt ni, semoga terus unggul dan semoga dengan terpilih lagi Indonesia lebih baik lagi #jokowi... https://t.co/koeBaHyZsC
14	selamat untuk Pak Jokowi, #jokowi #berSATUikutKyai https://t.co/CsWBylrDlo
15	semoga terus unggul sampai sudah 100% real count nya #jokowi #berSATUikutKyai https://t.co/Nep6NRM0DX
16	ini hasil real count dari KPU per 4 Mei kemaren #jokowi #berSATUikutKyai https://t.co/IJqsRuFOHt
17	Semalam kakakku bilang, "ambil cuti tgl 22 mei, kakak gak mau ada tambahan pengajian tahun depan" #22mei... https://t.co/jRvwqK2CEH
18	RT @pantau_com: PAN Merapat ke Jokowi, Amien Rais: Omong Kosong! https://t.co/jepe5uEPEq #pantaucom #partaiamanatnasional #amienrais #koali...
19	Jokowi Unggul di Pyongyang, Prabowo Menang di Karachi #jokowi #berSATUikutKyai https://t.co/lyH0Wegmf2 https://t.co/UqNDDAQssD
20	PAN Merapat ke Jokowi, Amien Rais: Omong Kosong! https://t.co/jepe5uEPEq #pantaucom #partaiamanatnasional #amienrais #koalisi #jokowi

Pada Tabel di atas dijelaskan bahwa dokumen yang diakuisisi adalah dokumen berlabel text media social twitter kemudian dokumen dimasukkan sebagai proses pembelajaran. Pada langkah ini data masih utuh atau belum dibersihkan sehingga data opini text yang akan di proses masih bercampur dengan karakter-karakter lainnya.

5.2. Preprocessing

5.2.1. Pembersihan Dokumen

Table 4 PEMBERSIHAN DOKUMEN

NO	TWEET
1	Dikit dikit tagar dikit dikit tagar, tagar kok dikit dikit
2	Sakdiyah Ma'ruf: Indonesia Penuh Lelucon Politik, Saatnya Komedi Memimpin
3	Ketegangan dalam bekerja akan menimbulkan dampak yg besar bagi kesehatan kita mgkn itu yg terjadi pada para pejuang...
4	Untuk alasan apapun, ketua MPR harus diganti dan biarkan PAN jadi oposisi sampai 2024.
5	One Belt One Road (OBOR), Modus Invansi Senyap Ala Cina Klik
6	Soal Peluang Demokrat Gabung KIK, Ferdinand: Tergantung Pak Jokowi
7	Prabowo 'Ngambek' karena AHY Penuhi Undangan Jokowi? Andi Arief Berkicau
8	Siapapun yang terpilih harus menjadi Kemenangan rakyat Indonesia.
9	Ketua DPC Gerindra Apresiasi Kinerja KPU dan Bawaslu Bangkalan
10	Grace Natalie Tak #Pantas Jadi Menteri Kabinet
11	selamat dan semoga Indonesia semakin Maju
12	mantap Pak Jokowi dan Kyai Ma'ruf, selamat ya
13	mantap bgt ni, semoga terus unggul dan semoga dengan terpilih lagi Indonesia lebih baik lagi
14	selamat untuk Pak Jokowi,
15	semoga terus unggul sampai sudah 100% real count nya
16	ini hasil real count dari KPU per 4 Mei kemaren
17	Semalam kakakku bilang, "ambil cuti tgl 22 mei, kakak gak mau ada tambahan pengajian tahun depan"
18	RT @pantau_com: PAN Merapat ke Jokowi, Amien Rais: Omong Kosong!
19	Jokowi Unggul di Pyongyang, Prabowo Menang di Karachi

20	PAN Merapat ke Jokowi, Amien Rais: Omong Kosong!
----	--

Pada Tabel di atas dijelaskan bahwa dokumen yang diakuisisi dibersihkan dari setiap karakter seperti <https://>, hastag, alamat site, username, tanda baca seperti (.,?![]/%;:<>()), angka(0,1,2,3,4,5,6,7,8,9) dan jugakarakter lain selain alphabet.

5.2.2. Case Folding

Pada proses ini dilakukan penyeragaman bentuk huruf dalam dokumen , dimana huruf kapital diubah menjadi huruf kecil dan huruf diseragamkan dari A sampai Z dan selain huruf akan dihilangkan karena dianggap delimiter.

Table 5 CASE FOLDING

NO	TWEET
1	dikit dikit tagar dikit dikit tagar tagar kok dikit dikit
2	sakdiyah maruf indonesia penuh lelucon politik, saatnya komedi memimpin
3	ketegangan dalam bekerja akan menimbulkan dampak yg besar bagi kesehatan kita mgkn itu yg terjadi pada para pejuang
4	untuk alasan apapun, ketua mpr harus diganti dan biarkan pan jadi oposisi sampai 2024
5	one belt one road obor modus invansi senyap ala cina klik
6	soal peluang demokrat gabung kik ferdinand tergantung pak jokowi
7	prabowo ngambek karena ahy penuhi undangan jokowi andi arief berkicau
8	siapapun yang terpilih harus menjadi kemenangan rakyat indonesia.
9	ketua dpc gerindra apresiasi kinerja kpu dan bawaslu bangkalan
10	grace natalie tak #pantas jadi menteri kabinet
11	selamat dan semoga indonesia semakin maju
12	mantap pak jokowi dan kyai ma'ruf, selamat ya

13	mantap bgt ni, semoga terus unggul dan semoga dengan terpilih lagi indonesia lebih baik lagi
14	selamat untuk pak jokowi,
15	semoga terus unggul sampai sudah 100 real count nya
16	ini hasil real count dari kpu per 4 mei kemaren
17	semalam kakakku bilang, ambil cuti tgl 22 mei, kakak gak mau ada tambahan pengajian tahun depan
18	pan merapat ke jokowi amien rais omong kosong!
19	jokowi unggul di pyongyang, prabowo menang di karachi
20	pan merapat ke jokowi, amien rais: omong kosong!

5.2.3. Parsing

Table 6 PARSING

A1	dikit	dikit	tagar	dikit	dikit	tagar	tagar	kok
A2	sakdiyah	maruf	indonesia	penuh	lelucon	politik	saatnya	komedi
A3	one	belt	one road	obor	modus	senyap	ala	cina
A4	siapapun	yang	terpilih	harus	menjadi	kemenangan	rakyat	indonesia
A5	grace	natalie	tak	jadi	mentri	kabinet		
A6	selamat	dan	sukses	indonesia	semakin	maju		
A7	mantap	pak	jokowi	dan	kyai	maruf	selamat	ya
A8	jokowi	unggul	di	pyongyang	prabowo	menang	di	karawaci

Pada tabel di atas dijelaskan bahwa langkah parsing atau tokenizing, pada tahapan ini kalimat di pecah2 menjadi kata-kata kemudian menganalisa terhadap kumpulan kata dengan memisahkan kata itu dan juga menentukan struktur sintaksis data tiap kata.

5.2.3. Filtering

Table 7 FILTERING

A1	dikit		tagar					kok
A2	sakdiyah	maruf	indonesia	penuh	lelucon	politik	saatnya	komedi
A3	one	belt	oneroad	obor	modus	senyap		cina
A4	siapapun		terpilih	harus	menjadi	kemenangan	rakyat	indonesia
A5	grace	natalie		jadi	mentri	kabinet		
A6	selamat		sukses	indonesia	semakin	maju		
A7	mantap	pak	jokowi		kyai	maruf	selamat	
A8	jokowi	unggul		pyongyang	prabowo	menang		karawaci

Pada tabel di atas dijelaskan tahapan menghilangkan kata yang tidak penting berdasarkan stopword seperti kata penghubung (di,ke,yang,ya) atau kata yang tidak mempunyai makna. Tahapan ini menyulitkan karena komentar orang saat ini tidak memakai bahasa Indonesia yang baku dan benar sehingga menyulitkan di dalam penghapusan data.

5.2.3. Stemming

Table 8 STEAMMING

A1	dikit	dikit	tagar	dikit	dikit			kok
A2	sakdiyah	maruf	indonesia	penuh	lelucon	politik	saat	komedi
A3	one	belt	oneroad	obor	modus	senyap	rakyat	cina
A4	siapa		terpilih	harus	jadi	menang	rakyat	indonesia
A5	grace	natalie		jadi	mentri	kabinet		
A6	selamat		sukses	indonesia	semakin	maju		
A7	mantap	pak	jokowi		kyai	maruf	selamat	
A8	jokowi	unggul		pyongyang	prabowo	menang		karawaci

Pada tahan stemming mengubah kata menjadi kata dasar dengan menghilangkan affixes yaitu awalan prefixes, sisipan infixes, akhiran suffixes dan confixes pada kata turunan yang ada pada dokumen dan di cocokan dengan KBBI dan di rubah sesuai aturan stemming.

5.2.3. Proses Pembobotan

Tahapan pembobotan ini untuk menghitung dan informasi term frequency, document frequency dan inverse dokument frequency, dimana menghitung dokumen atau term ini berdasarkan frekuensi kemunculan term dokumen. Term atau kata dihitung probabilitas kemunculannya dalam satu dokumen A1-A8. untuk mendapatkan Inverse dokument frekuensi dengan menggunakan persamaan.

Nilai Inverse dokumet frekuensi dapat di rumuskan dalam persamaan :

$$IDF = \log(D/DF)$$

Table 9 PROSES PEMBOBOTAN

No	Kata	Dokument				DF	IDF	TF*IDF			
		A1	A2	A3	A4			A1	A2	A3	A4
1	dikit	4	0	0	0	4	0,602	2,408	0,000	0,000	0,000
2	terpilih	0	0	0	1	1	0,602	0,000	0,000	0,000	0,602
3	indonesia	0	1	0	1	2	0,301	0,000	0,602	0,000	0,602
4	menang	0	0	0	1	1	0,602	0,000	0,000	0,000	0,602
5	rakyat	0	0	1	1	2	0,301	0,000	0,000	0,602	0,602
6	politik	0	0	1	0	1	0,602	0,000	0,000	0,602	0,000
7	lelucon	0	1	0	0	1	0,602	0,000	0,602	0,000	0,000
8	tagar	1	0	0	0	1	0,602	0,602	0,000	0,000	0,000

5.3. Klasifikasi Dengan Support Vector Mechine (SVM)

Table 10 PEMISAHAN DATA PELATIHAN DAN DATA PENGUJIAN

PERBANDINGAN DATA		JUMLAH DATA	
DATA PELATIHAN	DATA PENGUJIAN	DATA PELATIHAN	DATA PENGUJIAN
80%	20%	800	200
70%	30%	700	300
60%	40%	600	400
50%	50%	500	500
40%	60%	400	600
30%	70%	300	700
20%	80%	200	800

Pada tabel di atas didapatkan hasil akuisisi data lewat preprocessing data yang ada sebanyak 1000 dibagi 3 kelas sentiment yaitu positif, netral, dan negatif. Data yang telah selesai di normalisasi sebelum dimasukkan ke klasifikasi data di bagi 2 pelatihan dan data pengujian dengan menggunakan metode Cross Validation. pada pengujian ini dilakukan sebanyak 7 kali dengan data training dan testing yang berbeda-beda.

Untuk model pembelajaran diklasifikasikan dengan percobaan pengujian sebanyak 7 kali dimana setiap satu kali pengujian di dapatkan metrix dengan ukuran 3x3 sebagai representatif kelas aktual dan prediksi. Sesudah di dapatkan hasil dari SVM dilanjutkan dengan percobaan pengujian menggunakan data uji yang telah dipisahkan sebanyak 7 kali.

Table 11 CONFUSION MATRIC

Aktual Data	Prediksi		
	POSITIF	NETRAL	NEGATIF
POSITIF	TP	PN	FP
NETRAL	NP	TN	NNg
NEGATIF	FN	NgN	TNg

Untuk penjelasan istilah-istilah di table adalah sebagai berikut:

- TP : Kelas kata terprediksi benar bernilai positif
- FP : Kelas kata positif terprediksi negatif
- PN : Kelas kata positif terprediksi netral
- TN : Kelas kata terprediksi netral
- NP : Kelas kata netral terprediksi sebagai kelas kata positif
- NNg: Kelas kata netral terprediksi negatif
- NgN: Kelas Kata Negatif terprediksi netral
- FN : Kelas Kata negatif terprediksi kata positif
- TNg: Kelas Kata Negatif terprediksi Negatif

Pada tabel di atas dapat di jelaskan bahwa confussion matrix adalah hasil prediksi memakai SVM yang di ukur performa nya dari setiap kelas yaitu dengan cara menghitung presion dan recall.

5.3. Percobaan dengan data Pelatihan 80% dan Pengujian 20%

Table 12 HASIL DARI PERCOBAAN PERTAMA

Aktual Data	Prediksi			Precision	Recall	Kelas
	POSITIF	NETRAL	NEGATIF			
POSITIF	12	133	0	0,255	0,085	Positif
NETRAL	26	455	0	0,689	0,945	Netral
NEGATIF	9	72	0	0	0	Negatif

Pada tabel diatas dapat di lihat percobaan pertama digunakan dataset yang telah di pisahkan secara random dan di dapatkan data pelatihan 80% dan data pengujian 20% yang di klasifikasikan dengan SVM maka menghasilkan confusion matrix.

Pada file CSV hasil pengolahan filter, dilakukan klasifikasi sentimen untuk tiap kalimat dengan pemberian label positif, negatif, dan netral. Ujicoba menggunakan syntax python pada editor pycharm untuk penerapan metode.

Precision, recall dan accuracy merupakan parameter yang digunakan untuk mengetahui seberapa baik kualitas dari pengujian. Precision adalah tingkat kecocokan antara informasi yang dicari oleh pengguna dengan hasil yang diberikan sistem. *Recall* adalah tingkat kesuksesan sistem dalam mendapatkan informasi kembali. Dan *accuracy* adalah tingkat kedekatan antara nilai prediksi dengan nilai sebenarnya yang sudah diklasifikasi dengan benar.

Dari hasil nilai pada diagram menunjukkan bahwa hasil pengujian belum dianggap baik. Hal ini bisa terjadi karena banyak data opini positif tergolong sebagai negatif maupun sebaliknya. Selain itu, juga banyak data opini yang terklasifikasi netral

menjadi tergolong sebagai opini positif atau negatif. Berikut merupakan permasalahan yang menyebabkan hasil dari metode *Lexicon Based* belum dianggap baik.

Pada Lexicon base di temukan adanya kata ambigu, dan dalam melakukan penentuan, SentiWordNet memiliki pengertian kata yang banyak dan juga nilai sentimen yang berbeda. Satu kata pada SentiWordNet dapat mewakili lebih dari satu makna sehingga dengan metode yang dipilih dapat menimbulkan kesalahan pemilihan makna kata. Karena pada lexicon based dengan metode yang dipilih yaitu first sense pencocokan frase dengan data sentimen yang ada pada dictionary menjadi bersifat tunggal. Artinya satu kata hanya memiliki satu sentimen yaitu pada kata yang muncul paling atas atau paling populer . Adanya kesalahan ini juga berakibat pada salahnya nilai sentimen.