

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

BAB 4 PEMBAHASAN

4.1 Pengumpulan Data

Pengumpulan data untuk penelitian ini diambil melalui *crawling* data *Youtube* dengan menggunakan library python. Adapun hasil dari *crawling* tersebut mengumpulkan sebanyak 4629 komentar terhitung dari video di upload sampai dengan tanggal 16 Agustus 2023. Berikut ini hasil *crawling* data disajikan dalam Tabel 4.1

Tabel 4. 1 Hasil Crawling Dataset

No	Komentar	Sentimen
1	Harusnya pemerintahan perlu kandidat muda biar berpikinya gak kolot,biar pemikiran melenial terbuka,kalau gak negara ini gak akan maju maju	Negatif
2	Kami sangat setuju pembatalan tuan rumah u20.. karena komitmen bangsa kita tidak tergoyahkan apapun yang terjadi.	Positif
3	Sebenarnya sederhana palestina (yg dibela mati2an) sama sekali TIDAK mempermasalahkan israel ikut tanding,lalu kenapa indonesia malah mempermasalahkan?? Sampai harus mengorbankan impian garuda muda?	Netral

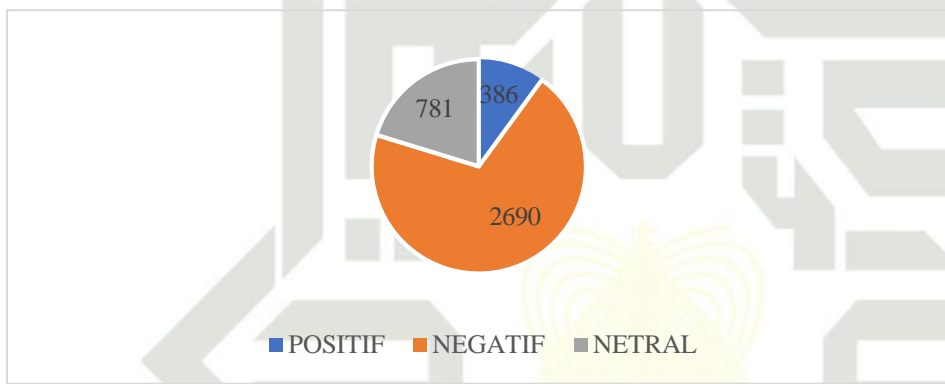
4.2 Labeling Text

Labeling atau pemberian label pada setiap komentar ini dilakukan bersama beberapa mahasiswa dengan menganalisa komentar tersebut apakah masuk ke dalam kelas positif, negatif, atau netral. Dalam menentukan kelas sentiimen pada dataset dilakukan beberapa pendekatan yaitu untuk komentar yang berisikan kalimat pujian, semangat dan rasa bangga di masukkan kedalam kategori kelas positif, sementara untuk kamentar seperti ujaran kebencian, rasa kecewa dan rasa

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

marah dimasukkan kedalam kategori negatif. Sedangkan untuk menentukan kelas adalah komentar yang tidak termasuk kedalam kategori positif dan negatif. Dari data yang dianalisa menghasilkan 4629 data kemudian dilakukan proses filtering dengan menghapus data duplikat dan data yang tidak sesuai dengan kasus sehingga data yang dihasilkan setelah proses filtering sebanyak 3857 data dengan jumlah data dengan sentimen negatif sebanyak 2690 data, sentimen positif 386 data dan sentimen netral 781 data. berikut ini disajikan diagram distribusi kelas pada dataset pada gambar 3.2



Gambar 3. 2 Distibusi Sentimen dalam Dataset

4.3 Preprocessing Teks

Pra-pemrosesan teks merupakan langkah awal di mana data teks yang mengandung gangguan akan diurutkan untuk memudahkan pengolahan lebih lanjut dan mengurangi gangguan tersebut. Dibawah ini disajikan hasil dari tahapan-tahapan pada pre-processing.

4.3.1 Cleaning

Dalam tahap cleaning, dilakukan penghapusan tanda baca, angka, simbol, dan URL, dan nama pengguna dari teks. Sementara itu, eliminasi stopwords bertujuan untuk menghilangkan kata-kata yang dianggap tidak relevan dalam teks. berikut di sajikan pada tabel dibawah ini.

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

Tabel 4. 2 Cleaning

Sebelum Cleaning	Setelah Cleaning
Ulama dan Santri dukung Ganjar Calon Presiden 2024 https://t.co/YjVKFi7evT	Ulama dan Santri dukung Ganjar Calon Presiden
cemberutnya mana pak gk diajak PDIP waktu penetapan ganjar Calon Presiden? katanya wilayah parpol?	cemberutnya mana pak gk diajak PDIP waktu penetapan ganjar Calon Presiden katanya wilayah parpol
@SINDOnews Siapapun cawapres nya , Calon Presiden nya tetep Pak Ganjar □ □	Siapapun cawapres nya Calon Presiden nya tetep Pak Ganjar

4.3.2 Case Folding

Case folding adalah tahapan pengubahan data menjadi huruf kecil. Berikut disajikan pada tabel dibawah ini.

Tabel 4. 3 Case Folding

Sebelum Casefolding	Setelah Casefolding
Ulama dan Santri dukung Ganjar Calon Presiden 2024 https://t.co/YjVKFi7evT	ulama dan santri dukung ganjar Calon Presiden
cemberutnya mana pak gk diajak PDIP waktu penetapan ganjar Calon Presiden? katanya wilayah parpol?	cemberutnya mana pak gk diajak pdip waktu penetapan ganjar Calon Presiden katanya wilayah parpol
@SINDOnews Siapapun cawapres nya , Calon Presiden nya tetep Pak Ganjar □ □	siapapun cawapres nya Calon Presiden nya tetep pak ganjar

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

4.3.3 Tokenizing

Tokenizing adalah tahapan penguraian komentar menjadi potongan kata. Berikut disajikan pada tabel dibawah ini.

Tabel 4. 4 Tokenizing

Sebelum Tokenizing	Setelah Tokenizing
Ulama dan Santri dukung Ganjar Calon Presiden 2024 https://t.co/YjVKFi7evT	['ulama', 'dan', 'santri', 'dukung', 'ganjar', 'Calon Presiden', '']
cemberutnya mana pak gk diajak PDIP waktu penetapan ganjar Calon Presiden? katanya wilayah parpol?	['cemberutnya', 'mana', 'pak', 'tidak', 'diajak', 'pdip', 'waktu', 'penetapan', 'ganjar', 'Calon Presiden', 'katanya', 'wilayah', 'parpol', '']
@SINDOnews Siapapun cawapres nya ,Calon Presiden nya tetep Pak Ganjar□□	['', 'siapapun', 'cawapres', 'nya', 'Calon Presiden', 'nya', 'tetap', 'pak', 'ganjar']

4.4 Normalisasi

Normalisasi merubah jenis bahasa yang tidak resmi (slang words) atau kata yang tidak baku menjadi kata baku. Berikut disajikan pada tabel dibawah ini.

Tabel 4. 5 Normalisasi

Sebelum Normalisasi	Setelah Normalisasi
Ulama dan Santri dukung Ganjar Calon Presiden 2024 https://t.co/YjVKFi7evT	['ulama', 'dan', 'santri', 'dukung', 'ganjar', 'Calon Presiden', '']
cemberutnya mana pak gk diajak PDIP waktu penetapan ganjar Calon	['cemberutnya', 'mana', 'pak', 'tidak', 'diajak', 'pdip', 'waktu', 'penetapan', 'ganjar', 'Calon Presiden', 'katanya',



Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

Presiden? katanya wilayah parpol?	'wilayah', 'parpol', ''
@SINDOnews Siapapun cawapres nya , Calon Presiden nya tetep Pak Ganjar□□	['', 'siapapun', 'cawapres', 'nya', 'Calon Presiden', 'nya', 'tetap', 'pak', 'ganjar']

4.3.5 Stopword Removal

Stopword removal adalah tahapan menghilangkan kata penghubung dalam sebuah kalimat, misalnya “dan”, “apa” dan lain sebagainya. Berikut disajikan pada tabel dibawah ini.

Tabel 4. 6 Stopword Removal

Sebelum Stopword Removal	Setelah Stopword Removal
Ulama dan Santri dukung Ganjar Calon Presiden 2024 https://t.co/YjVKFi7evT	['ulama', '', 'santri', 'dukung', '', 'Calon Presiden', '']
cemberutnya mana pak gk diajak PDIP waktu penetapan ganjar Calon Presiden? katanya wilayah parpol?	['cemberutnya', '', '', 'diajak', 'pdip', 'waktu', 'penetapan', '', 'Calon Presiden', '', 'wilayah', 'parpol', '']
@SINDOnews Siapapun cawapres nya , Calon Presiden nya tetep Pak Ganjar □□	['', '', 'cawapres', '', 'Calon Presiden', '', '', '', '']

4.3.6 Stemming

Stemming adalah proses menghilangkan imbuhan, sisipan dan akhiran dalam sebuah kata. Berikut disajikan pada tabel dibawah ini.

Tabel 4. 7 Stemming

Sebelum Stemming	Setelah Stemming
------------------	------------------

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

Ulama dan Santri dukung Ganjar Calon Presiden 2024 https://t.co/YjVKFi7evT	['ulama', ', ', 'santri', 'dukung', ', ', 'Calon Presiden', ', ']
cemberutnya mana pak gk diajak PDIP waktu penetapan ganjar Calon Presiden? katanya wilayah parpol?	['cemberut', ', ', ', ', 'ajak', 'pdip', 'waktu', 'tetap', ', ', 'Calon Presiden', ', ', 'wilayah', 'parpol', ', ']
@SINDOnews Siapapun cawapres nya ,Calon Presiden nya tetep Pak Ganjar□□	['', ', ', 'cawapres', ', ', 'Calon Presiden', ', ', ', ', ', '']

4.4 Transformation Teks

Proses mengubah kata menjadi angka atau vektor dikenal sebagai transformasi teks. Dan frekuensi term yang muncul dalam dokumen yang relevan disebut TF (Term Frequency). Nilai kesesuaian atau bobot suatu term berkorelasi positif dengan jumlah kali term tersebut muncul dalam dokumen.

Namun, TF-IDF adalah teknik untuk menentukan nilai term frekuensi (tf) dan invers dokumen frekuensi (IDF) pada setiap kata (token) di setiap dokumen dalam korpus.

Dalam metode ini, akan dilakukan perhitungan bobot untuk setiap kata dalam dokumen menggunakan rumus:

$$W_{dt} = t_{fdt} * IDF_t$$

Dimana:

d : dokumen ke-d

t : kata ke-t dari kata kunci

W : bobot dokumen ke-d terhadap kata ke-t

tf : banyak kata yang dicari pada sebuah dokumen

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

IDF : *inversed document frequenc*

Untuk nilai IDF itu sendiri didapatkan dari **IDF : $\log_2(D/df)$**

Dimana:

D : total dokumen

df : banyak dokumen yang mengandung kata yang banyak dicari

1. Langkah awal adalah menghitung Term Frequency (TF), yang dilakukan dengan memberikan nilai 1 atau 0 pada setiap term dalam teks. Nilai-nilai ini kemudian direpresentasikan dalam tabel di bawah ini.

Tabel 4. 8 Hasil Pembobotan TF

Kata(Term)	TF		
	D1	D2	D3
ulama	1	0	0
santri	1	0	0
dukung	1	0	0
Calon Presiden	1	0	0
cemberut	0	1	0
ajak	0	1	0
pdip	0	1	0
waktu	0	1	0
tetap	0	1	0
Calon Presiden	0	1	0
wilayah	0	1	0

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

parpol	0	1	0
cawapres	0	0	1
Calon Presiden	0	0	1

2. Langkah kedua adalah menemukan DF dengan menjumlahkan bobot nilai term dalam dokumen. seperti yang disajikan pada tabel di bawah ini.

Tabel 4. 9 Hasil Pembobotan DF

Kata(Term)	TF			IDF
	D1	D2	D3	TF
ulama	1	0	0	1
santri	1	0	0	1
dukung	1	0	0	1
Calon Presiden	1	0	0	1
cemberut	0	1	0	1
ajak	0	1	0	1
pdip	0	1	0	1
waktu	0	1	0	1
tetap	0	1	0	1
Calon Presiden	0	1	0	1
wilayah	0	1	0	1
parpol	0	1	0	1
cawapres	0	0	1	1

Calon Presiden	0	0	1	1
----------------	---	---	---	---

3. Langkah ketiga menghitung IDF dengan menggunakan rumus:

IDF : $\log_2(D/df)$.

Dimana :

D : total dokumen

Df : banyak dokumen yang mengandung kata yang banyak dicari

Tabel 4. 10 Hasil Pembobotan IDF

Kata(Term)	TF			IDF	
	D1	D2	D3	TF	$\log_2(D/df)$
ulama	1	0	0	1	0,301029996
santri	1	0	0	1	0,301029996
dukung	1	0	0	1	0,301029996
Calon Presiden	1	0	0	1	0,301029996
cemberut	0	1	0	1	0,301029996
ajak	0	1	0	1	0,301029996
pdip	0	1	0	1	0,301029996
waktu	0	1	0	1	0,301029996
tetap	0	1	0	1	0,301029996
Calon Presiden	0	1	0	1	0,301029996
wilayah	0	1	0	1	0,301029996
parpol	0	1	0	1	0,301029996

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

cawapres	0	0	1	1	0,301029996
Calon Presiden	0	0	1	1	0,301029996

4. Langkah keempat menghitung bobot TF-IDF.

Pada tabel 4.12 menunjukkan hasil pembobotan yang diperoleh berdasarkan data komentar yang ditemukan di tabel sebelumnya. Bobot akhir suatu term adalah mengalikan keduanya, yaitu TF x IDF.

Tabel 4. 11 Hasil Pembobotan TF-IDF

Kata(Term)	TF			IDF				
	D1	D2	D3	TF	$\log_2(D/d)$ f)	D1	D2	D3
ulama	1	0	0	1	0,301029 996	0,301029 996	0	0
santri	1	0	0	1	0,301029 996	0,301029 996	0	0
dukung	1	0	0	1	0,301029 996	0,301029 996	0	0
Calon Presiden	1	0	0	1	0,301029 996	0,301029 996	0	0
cemberut	0	1	0	1	0,301029 996	0	0,301029 996	
ajak	0	1	0	1	0,301029 996	0	0,301029 996	0
pdip	0	1	0	1	0,301029 996	0	0,301029 996	0
waktu	0	1	0	1	0,301029	0	0,301029	0



Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

					996		996	
tetap	0	1	0	1	0,301029 996	0	0,301029 996	0
Calon Preside n	0	1	0	1	0,301029 996	0	0,301029 996	0
wilaya h	0	1	0	1	0,301029 996	0	0,301029 996	0
parpol	0	1	0	1	0,301029 996	0	0,301029 996	0
cawapr es	0	0	1	1	0,301029 996	0	0	0,3010 29996
Calon Preside n	0	0	1	1	0,301029 996	0	0	0,3010 29996

4.5 Proses Klasifikasi Naïve Bayes

Pada proses ini, akan diberikan penjelasan tentang metode yang akan digunakan untuk mengklasifikasikan data.

Seluruh kumpulan data dibagi menjadi beberapa langkah data latih dan data uji selama proses klasifikasi Naive Bayes ini. Berikut adalah penjelasan prosedurnya dan contoh perhitungan:

1. Tahap Data Latih :

Pada proses ini, nilai bobot TF-IDF setiap kata akan digunakan sebagai referensi untuk menghitung nilai kemungkinan sebelumnya untuk setiap kategori, serta kemungkinan untuk setiap kata pada setiap term untuk setiap kelas data latih. Berikut adalah hasilnya:



Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

- a. Pertama hitung probabilitas *prior* setiap kategori. pada penelitian ini yang menjadi kategori ada tiga yaitu kategori positif, negatif dan netral.

$$P(\text{Positif/negatif/netral}) = \frac{D(\text{positif/negatif/netral})}{|C|}$$

Keterangan :

$p(\text{pos/neg/net})$ = Peluang kemunculan dokumen dari kelas positif, negatif dan netral

$d(\text{pos/neg/net})$ = Jumlah seluruh dokumen pada setiap kelas

$|c|$ = Jumlah keseluruhan data yang dilatih yaitu positif, negatif dan netral

$$P(\text{Pos}) = \frac{d(\text{Pos})}{|c|} = \frac{1}{3} = 0,3333333333$$

$$P(\text{Neg}) = \frac{d(\text{Neg})}{|c|} = \frac{1}{3} = 0,3333333333$$

$$P(\text{Net}) = \frac{d(\text{Net})}{|c|} = \frac{1}{3} = 0,3333333333$$

- b. Setelah melakukan perhitungan probabilitas setiap kategori kemudian selanjutnya menghitung probabilitas *likelihood* pada setiap *term* dari semua dokumen. Banyak *term* tergantung pada hasil preproses data. Berikut adalah merupakan perhitungan probabilitas pada setiap *term*:

$$P(w | (\text{pos/neg})/\text{net}) = \frac{(nk | (\text{pos/neg})/\text{net}) + 1}{(n, |(\text{pos/neg})/\text{net}) + |\text{kosakata}|}$$

Keterangan :

$p(w_k | \text{pos/neg/net})$ = Peluang kemunculan kata pada kategori

w_k = Kata yang muncul pada sebuah kategori.

$(nk, \text{pos/neg/net}) + 1$ = Jumlah frekuensi kemunculan kata pada kategori.



Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

n_k

= Kemunculan setiap kata pada kategori.

$|kosakata|$

= Jumlah semua kata dari semua kategori.

Dibawah ini terdapat perhitungan probabilitas kemunculan kata dalam setiap kelas sesuai dengan hasil tahapan *stemming* yang sudah dilakukan.

Diketahui :

$|kata| = 14$

Positif = 4

Negatif = 8

Netral = 2

1) Probabilitas kata “ulama”

$$p(\text{ulama}|\text{pos}) \frac{(\text{ulama}|\text{pos})+1}{(\text{pos}) + |kosakata|} = \frac{0,301029996 + 1}{4 + 14} = \mathbf{0,0722794442}$$

$$p(\text{ulama}|\text{neg}) \frac{(\text{ulama}|\text{neg})+1}{(\text{neg}) + |kosakata|} = \frac{0 + 1}{8 + 14} = \mathbf{0,0454545455}$$

$$p(\text{ulama}|\text{net}) \frac{(\text{ulama}|\text{net})+1}{(\text{net}) + |kosakata|} = \frac{0 + 1}{2 + 14} = \mathbf{0,0625}$$

2) Probabilitas kata “santri”

$$p(\text{santri}|\text{pos}) \frac{(\text{santri}|\text{pos})+1}{(\text{pos}) + |kosakata|} = \frac{0,301029996 + 1}{4 + 14} = \mathbf{0,0722794442}$$

$$p(\text{santri}|\text{neg}) \frac{(\text{santri}|\text{neg})+1}{(\text{neg}) + |kosakata|} = \frac{0 + 1}{8 + 14} = \mathbf{0,0454545455}$$

$$p(\text{santri}|\text{net}) \frac{(\text{santri}|\text{net})+1}{(\text{net}) + |kosakata|} = \frac{0 + 1}{2 + 14} = \mathbf{0,0625}$$

3) Probabilitas kata “dukung”



Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

$$p(\text{dukung}|\text{pos}) \frac{(\text{dukung}|\text{pos})+1}{(\text{pos}) + |\text{kosakata}|} = \frac{0,301029996 + 1}{4 + 14} = \mathbf{0,0722794442}$$

$$p(\text{dukung}|\text{neg}) \frac{(\text{dukung}|\text{neg})+1}{(\text{neg}) + |\text{kosakata}|} = \frac{0 + 1}{8 + 14} = \mathbf{0,0454545455}$$

$$p(\text{dukung}|\text{net}) \frac{(\text{dukung}|\text{net})+1}{(\text{net}) + |\text{kosakata}|} = \frac{0 + 1}{2 + 14} = \mathbf{0,0625}$$

4) Probabilitas kata “Calon Presiden”

$$p(\text{Calon Presiden}|\text{pos}) \frac{(\text{capres}|\text{pos})+1}{(\text{pos}) + |\text{kosakata}|} = \frac{0,301029996 + 1}{4 + 14} = \mathbf{0,0722794442}$$

$$p(\text{Calon Presiden}|\text{neg}) \frac{(\text{capres}|\text{neg})+1}{(\text{neg}) + |\text{kosakata}|} = \frac{0 + 1}{8 + 14} = \mathbf{0,0454545455}$$

$$p(\text{Calon Presiden}|\text{net}) \frac{(\text{capres}|\text{net})+1}{(\text{net}) + |\text{kosakata}|} = \frac{0 + 1}{2 + 14} = \mathbf{0,0625}$$

5) Probabilitas kata “cemberut”

$$p(\text{cemberut}|\text{pos}) \frac{(\text{cemberut}|\text{pos})+1}{(\text{pos}) + |\text{kosakata}|} = \frac{0 + 1}{4 + 14} = \mathbf{0,0555555556}$$

$$p(\text{cemberut}|\text{neg}) \frac{(\text{cemberut}|\text{neg})+1}{(\text{neg}) + |\text{kosakata}|} = \frac{0,301029996 + 1}{8 + 14} = \mathbf{0,0591377271}$$

$$p(\text{cemberut}|\text{net}) \frac{(\text{cemberut}|\text{net})+1}{(\text{net}) + |\text{kosakata}|} = \frac{0 + 1}{2 + 14} = \mathbf{0,0625}$$

6) Probabilitas kata “ajak”

$$p(\text{cemberut}|\text{pos}) \frac{(\text{cemberut}|\text{pos})+1}{(\text{pos}) + |\text{kosakata}|} = \frac{0 + 1}{4 + 14} = \mathbf{0,0555555556}$$

$$p(\text{cemberut}|\text{neg}) \frac{(\text{cemberut}|\text{neg})+1}{(\text{neg}) + |\text{kosakata}|} = \frac{0,301029996 + 1}{8 + 14} = \mathbf{0,0591377271}$$



Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengemukakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

$$p(\text{cemberut}|\text{net}) = \frac{(\text{cemberut}|\text{net})+1}{(\text{net}) + |\text{kosakata}|} = \frac{0 + 1}{2 + 14} = \mathbf{0,0625}$$

7) Probabilitas kata “pdip”

$$p(\text{pdip}|\text{pos}) = \frac{(\text{pdip}|\text{pos})+1}{(\text{pos}) + |\text{kosakata}|} = \frac{0 + 1}{4 + 14} = \mathbf{0,055555556}$$

$$p(\text{pdip}|\text{neg}) = \frac{(\text{pdip}|\text{neg})+1}{(\text{neg}) + |\text{kosakata}|} = \frac{0,301029996 + 1}{8 + 14} = \mathbf{0,0591377271}$$

$$p(\text{pdip}|\text{net}) = \frac{(\text{pdip}|\text{net})+1}{(\text{net}) + |\text{kosakata}|} = \frac{0 + 1}{2 + 14} = \mathbf{0,0625}$$

8) Probabilitas kata “waktu”

$$p(\text{waktu}|\text{pos}) = \frac{(\text{waktu}|\text{pos})+1}{(\text{pos}) + |\text{kosakata}|} = \frac{0 + 1}{4 + 14} = \mathbf{0,055555556}$$

$$p(\text{waktu}|\text{neg}) = \frac{(\text{waktu}|\text{neg})+1}{(\text{neg}) + |\text{kosakata}|} = \frac{0,301029996 + 1}{8 + 14} = \mathbf{0,0591377271}$$

$$p(\text{waktu}|\text{net}) = \frac{(\text{waktu}|\text{net})+1}{(\text{net}) + |\text{kosakata}|} = \frac{0 + 1}{2 + 14} = \mathbf{0,0625}$$

9) Probabilitas kata “tetap”

$$p(\text{tetap}|\text{pos}) = \frac{(\text{tetap}|\text{pos})+1}{(\text{pos}) + |\text{kosakata}|} = \frac{0 + 1}{4 + 14} = \mathbf{0,055555556}$$

$$p(\text{tetap}|\text{neg}) = \frac{(\text{tetap}|\text{neg})+1}{(\text{neg}) + |\text{kosakata}|} = \frac{0,301029996 + 1}{8 + 14} = \mathbf{0,0591377271}$$

$$p(\text{tetap}|\text{net}) = \frac{(\text{tetap}|\text{net})+1}{(\text{net}) + |\text{kosakata}|} = \frac{0 + 1}{2 + 14} = \mathbf{0,0625}$$

10) Probabilitas kata “Calon Presiden”

$$p(\text{capres}|\text{pos}) = \frac{(\text{capres}|\text{pos})+1}{(\text{pos}) + |\text{kosakata}|} = \frac{0 + 1}{4 + 14} = \mathbf{0,055555556}$$

$$p(\text{capres}|\text{neg}) = \frac{(\text{capres}|\text{neg})+1}{(\text{neg}) + |\text{kosakata}|} = \frac{0,301029996 + 1}{8 + 14} = \mathbf{0,0591377271}$$

$$p(\text{capres}|\text{net}) = \frac{(\text{capres}|\text{net})+1}{(\text{net}) + |\text{kosakata}|} = \frac{0 + 1}{2 + 14} = \mathbf{0,0625}$$



Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

11) Probabilitas kata “wilayah”

$$p(\text{wilayah}|\text{pos}) \frac{(\text{wilayah}|\text{pos})+1}{(\text{pos}) + |\text{kosakata}|} = \frac{0 + 1}{4 + 14} = \mathbf{0,0555555556}$$

$$p(\text{wilayah}|\text{neg}) \frac{(\text{wilayah}|\text{neg})+1}{(\text{neg}) + |\text{kosakata}|} = \frac{0,301029996 + 1}{8 + 14} = \mathbf{0,0591377271}$$

$$p(\text{wilayah}|\text{net}) \frac{(\text{wilayah}|\text{net})+1}{(\text{net}) + |\text{kosakata}|} = \frac{0 + 1}{2 + 14} = \mathbf{0,0625}$$

12) Probabilitas kata “parpol”

$$p(\text{parpol}|\text{pos}) \frac{(\text{parpol}|\text{pos})+1}{(\text{pos}) + |\text{kosakata}|} = \frac{0 + 1}{4 + 14} = \mathbf{0,0555555556}$$

$$p(\text{parpol}|\text{neg}) \frac{(\text{parpol}|\text{neg})+1}{(\text{neg}) + |\text{kosakata}|} = \frac{0,301029996 + 1}{8 + 14} = \mathbf{0,0591377271}$$

$$p(\text{parpol}|\text{net}) \frac{(\text{parpol}|\text{net})+1}{(\text{net}) + |\text{kosakata}|} = \frac{0 + 1}{2 + 14} = \mathbf{0,0625}$$

13) Probabilitas kata “cawapres”

$$p(\text{cawapres}|\text{pos}) \frac{(\text{cawapres}|\text{pos})+1}{(\text{pos}) + |\text{kosakata}|} = \frac{0 + 1}{4 + 14} = \mathbf{0,0555555556}$$

$$p(\text{cawapres}|\text{neg}) \frac{(\text{cawapres}|\text{neg})+1}{(\text{neg}) + |\text{kosakata}|} = \frac{0 + 1}{8 + 14} = \mathbf{0,0591377271}$$

$$p(\text{cawapres}|\text{net}) \frac{(\text{cawapres}|\text{net})+1}{(\text{net}) + |\text{kosakata}|} = \frac{0,301029996 + 1}{2 + 14} = \mathbf{0,0813143748}$$

14) Probabilitas kata “Calon Presiden”

$$p(\text{Calon Presiden}|\text{pos}) \frac{(\text{capres}|\text{pos})+1}{(\text{pos}) + |\text{kosakata}|} = \frac{0 + 1}{4 + 14} = \mathbf{0,0555555556}$$

$$p(\text{Calon Presiden}|\text{neg}) \frac{(\text{capres}|\text{neg})+1}{(\text{neg}) + |\text{kosakata}|} = \frac{0 + 1}{8 + 14} = \mathbf{0,0454545455}$$

$$p(\text{pdip}|\text{net}) \frac{(\text{pdip}|\text{net})+1}{(\text{net}) + |\text{kosakata}|} = \frac{0,301029996 + 1}{2 + 14} = \mathbf{0,0813143748}$$

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

2. Tahap Testing

Pada fase ini, langkah selanjutnya setelah menyelesaikan proses pelatihan adalah melakukan pengujian menggunakan data uji pada model yang telah dibuat sebelumnya. Sekali lagi dalam tahap ini, model akan dievaluasi dengan menginputkan data uji untuk melihat sejauh mana kinerja dan keefektifan model yang telah dilatih sebelumnya.

Tabel 4. 12 Data Testing

	Tweet	Kelas
1	Masyarakat setuju untuk mendukung Ganjar Pranowo sebagai Calon Presiden 2024 https://t.co/UrOoXWo1Hp	Positif
2	@Miduk17 Iya paket lengkap kalo ini, bahkan lebih pantas mahfud yg jadi Calon Presiden dari pada ganjar.	Negatif
3	aku jadi bayangin kalau pak Ganjar jadi presiden dan Pak Prabowo jadi Calon Presiden nya nih	Netral

Langkah berikutnya dalam proses ini adalah menghitung nilai probabilitas berdasarkan nilai setiap istilah. Perhitungan nilai probabilitas dilakukan dengan mengalikan probabilitas untuk setiap kategori dengan probabilitas setiap istilah yang diambil dari seluruh data. Sebelum melanjutkan ke tahap perhitungan (pengujian), data pengujian harus melalui serangkaian tahapan pra-pemrosesan terlebih dahulu. Berikut adalah hasil dari tahapan pra-pemrosesan untuk data pengujian.

Tabel 4. 13 Data Setelah Preprocessing

	Tweet	Kelas
--	--------------	--------------

masyarakat setuju untuk dukung ganjar pranowo sebagai Calon Presiden	Positif
paket lengkap mahfud Calon Presiden	Negatif
bayang ganjar presiden	Netral

Proses selanjutnya akan menghitung kemungkinan masing-masing kata dalam data uji dengan menggunakan hasil probabilitas term dari data latihan.

Tabel 4. 14 Data latihan dan uji yang sudah di preprocessing

Data Latih	Tweet	Kelas
	ulama santri dukung Calon Presiden	Positif
	cemberut ajak pdip waktu tetap Calon Presiden wilayah parpol	Negatif
	cawapres Calon Presiden	Netral
Data Uji	Tweet	Kelas
	masyarakat setuju untuk dukung ganjar pranowo sebagai Calon Presiden	?
	paket lengkap mahfud Calon Presiden	
	bayang ganjar presiden	

Tahapan berikutnya yaitu melakukan pembobotan kata pada data uji dapat dilihat pada tabel 4.16

- Hak Cipta Dilindungi Undang-Undang**
1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
 2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

Tabel 4. 15 Pembobotan Data Uji

Kata(Term)	TF			IDF		D1	D2	D3
	D1	D2	D3	TF	$\log_2(D/d)$			
masyarakat	1			1	0,301029 996	0,301029 996	0	0
setuju	1	0	0	1	0,301029 996	0,301029 996	0	0
dukung	1	0	0	1	0,301029 996	0,301029 996	0	0
ganjar	1	0	0	1	0,301029 996	0,301029 996	0	0
sebagai	1	0	0	1	0,301029 996	0,301029 996	0	0
Calon Presiden	1	0	0	1	0,301029 996	0,301029 996	0	0
lengkap	0	1	0	1	0,301029 996	0	0,301029 996	0
mahfud	0	1	0	1	0,301029 996	0	0,301029 996	0
ganjar	0	0	1	1	0,301029 996	0	0	0,3010 29996
presiden	0	0	1	1	0,301029 996	0	0	0,3010 29996

Diketahui :

Total nilai TF |kosakata|=10

Total Nilai Positif = 6



Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

Total Nilai Negatif = 2

Total Nilai Netral = 2

Selanjutnya dilakukan perhitungan probabilitas masing-masing kata diproses dengan menggunakan rumus dibawah ini.

$$p(wk | (pos/neg)/net) = \frac{(nk, pos/neg/net) + 1}{(n, pos/neg/net) + |kosakata|}$$

Probabilitas Data Uji (Positif)

$$p(\text{"masyarakat"} | \text{"pos"}) = (0.301029996 + 1) / (6 + 10) = 0,0813143748$$

$$p(\text{"setuju"} | \text{"pos"}) = (0.301029996 + 1) / (6 + 10) = 0,0813143748$$

$$p(\text{"dukung"} | \text{"pos"}) = (0.301029996 + 1) / (6 + 10) = 0,08131437481$$

$$p(\text{"ganjar"} | \text{"pos"}) = (0.301029996 + 1) / (6 + 10) = 0,0813143748$$

$$p(\text{"sebagai"} | \text{"pos"}) = (0.301029996 + 1) / (6 + 10) = 0,08131437481$$

$$p(\text{"Calon Presiden"} | \text{"pos"}) = (0.301029996 + 1) / (6 + 10) = 0,0813143748$$

$$p(\text{"lengkap"} | \text{"pos"}) = (0 + 1) / (6 + 10) = 0,0625$$

$$p(\text{"mahfud"} | \text{"pos"}) = (0 + 1) / (6 + 10) = 0,0625$$

$$p(\text{"ganjar"} | \text{"pos"}) = (0 + 1) / (6 + 10) = 0,0625$$

$$p(\text{"presiden"} | \text{"pos"}) = (0 + 1) / (6 + 10) = 0,0625$$

Probabilitas Data Uji (Negatif)

$$p(\text{"masyarakat"} | \text{"neg"}) = (0 + 1) / (2 + 10) = 0,0834$$

$$p(\text{"setuju"} | \text{"neg"}) = (0 + 1) / (2 + 10) = 0,0834$$

$$p(\text{"dukung"} | \text{"neg"}) = (0 + 1) / (2 + 10) = 0,0834$$

$$p(\text{"ganjar"} | \text{"neg"}) = (0 + 1) / (2 + 10) = 0,0834$$

$$p(\text{"sebagai"} | \text{"neg"}) = (0 + 1) / (2 + 10) = 0,0834$$

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

$$p(\text{"Calon Presiden"}|\text{"neg"}) = (0 + 1)/(2 + 10) = 0,0834$$

$$p(\text{"lengkap"}|\text{"neg"}) = (0.301029996 + 1)/(2 + 10) = 0,1084191663$$

$$p(\text{"mahfud"}|\text{"neg"}) = (0.301029996 + 1)/(2 + 10) = 0,1084191663$$

$$p(\text{"ganjar"}|\text{"neg"}) = (0 + 1)/(2 + 10) = 0,0834$$

$$p(\text{"presiden"}|\text{"neg"}) = (0 + 1)/(2 + 10) = 0,0834$$

Probabilitas Data Uji (Netral)

$$p(\text{"masyarakat"}|\text{"net"}) = (0 + 1)/(2 + 10) = 0,0834$$

$$p(\text{"setuju"}|\text{"net"}) = (0 + 1)/(2 + 10) = 0,0834$$

$$p(\text{"dukung"}|\text{"net"}) = (0 + 1)/(2 + 10) = 0,0834$$

$$p(\text{"ganjar"}|\text{"net"}) = (0 + 1)/(2 + 10) = 0,0834$$

$$p(\text{"sebagai"}|\text{"net"}) = (0 + 1)/(2 + 10) = 0,0834$$

$$p(\text{"Calon Presiden"}|\text{"net"}) = (0 + 1)/(2 + 10) = 0,0834$$

$$p(\text{"lengkap"}|\text{"net"}) = (0 + 1)/(2 + 10) = 0,0834$$

$$p(\text{"mahfud"}|\text{"net"}) = (0 + 1)/(2 + 10) = 0,0834$$

$$p(\text{"ganjar"}|\text{"net"}) = (0.301029996 + 1)/(2 + 10) = 0,1084191663$$

$$p(\text{"presiden"}|\text{"net"}) = (0.301029996 + 1)/(2 + 10) = 0,1084191663$$

Tahapan selanjutnya menghitung probabilitas kategori dengan rumus berikut :

$$p(\text{pos/neg}|d) = p(\text{pos/neg}) * \prod_i P(a_i|\text{pos/neg})$$

Keterangan :

$p(\text{pos/neg})$ = Peluang kemunculan kata pada kategori atau kelas dengan dokumen data uji.

d = Dokumen data baru

$\prod_i P(a_i|\text{pos/neg})$ = Peluang kemunculan a_i pada kategori atau kelas



Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

= Kata baru yang akan diuji.

Probabilitas kategori positif

$$P(uji|pos) = p(pos) \times p(masyarakat|pos) \times p(setuju|pos) \times p(dukung|pos) \times p(ganjar|pos) \times p(sebagai|pos) \times p(Calon Presiden|pos)$$

$$p(uji|pos) = 0,3333333333 \times 0,0813143748 \times 0,0813143748 \times 0,0813143748 \times 0,0813143748 \times 0,0813143748 \times 0,0813143748$$

$$= 0.000202399$$

$$P(uji|pos) = p(pos) \times p(lengkap|pos) \times p(mahfud|pos)$$

$$p(uji|pos) = 0,3333333333 \times 0,0625 \times 0,0625$$

$$= 0.00026041666$$

$$P(uji|pos) = p(pos) \times p(ganjar|pos) \times p(presiden|pos)$$

$$p(uji|pos) = 0,3333333333 \times 0,0625 \times 0,0625$$

$$= 0.00026041666$$

Probabilitas kategori negatif

$$P(uji|neg) = p(neg) \times p(masyarakat| neg) \times p(setuju|neg) \times p(dukung|neg) \times p(ganjar|neg) \times p(sebagai|neg) \times p(Calon Presiden|neg)$$

$$p(uji|neg) = 0,3333333333 \times 0,0834 \times 0,0834 \times 0,0834 \times 0,0834 \times 0,0834 \times 0,0834$$

$$= 0.00024301713$$

$$P(uji|neg) = p(neg) \times p(lengkap|neg) \times p(mahfud neg)$$

$$p(uji| neg) = 0,3333333333 \times 0,1084191663 \times 0,1084191663$$

$$= 0.0036905027$$

$$P(uji|neg) = p(neg) \times p(ganjar|neg) \times p(presiden|neg)$$

$$p(uji|neg) = 0,3333333333 \times 0,0834 \times 0,0834$$

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

$$= 0.0023119855$$

Probabilitas kategori netral

$$P(uji|net) = p(net) \times p(masyarakat|net) \times p(setuju|net) \times p(dukung|net) \times p(ganjar|net) \times p(sebagai|net) \times p(Calon Presiden|net)$$

$$p(uji|net) = 0,3333333333 \times 0,0834 \times 0,0834 \times 0,0834 \times 0,0834 \times 0,0834 \times 0,0834$$

$$= 0.00024301713$$

$$P(uji|net) = p(net) \times p(lengkap|net) \times p(mahfud net)$$

$$p(uji|net) = 0,3333333333 \times 0,0834 \times 0,0834$$

$$= 0.0023119855$$

$$P(uji|net) = p(net) \times p(ganjar|net) \times p(presiden|net)$$

$$p(uji|net) = 0,3333333333 \times 0,1084191663 \times 0,1084191663$$

$$= 0.0036905027$$

Berdasarkan hasil perhitungan di atas, nilai probabilitas tertinggi untuk kategori positif adalah 0.00026041666, sehingga dapat disimpulkan bahwa komentar tersebut tergolong ke dalam kelas "Positif". Di sisi lain, nilai probabilitas tertinggi untuk kategori negatif mencapai 0.0036905027, sehingga komentar tersebut dapat diklasifikasikan sebagai kelas "Negatif". Selain itu, nilai probabilitas tertinggi untuk kategori netral juga sama dengan 0.0036905027, sehingga komentar tersebut dinyatakan terklasifikasi sebagai kelas "Netral".

4.6 Proses Klasifikasi Naïve Bayes Menggunakan Python

Pada tahap pembahasan ini, akan dijelaskan serangkaian langkah dalam perancangan pemodelan. Proses pemodelan dimulai dengan memilih perpustakaan (library) yang akan digunakan untuk melakukan pra-pemrosesan teks, transformasi teks, menerapkan algoritma klasifikasi Naive Bayes, dan menguji teks yang telah diproses.

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

4.6.1 Library

Library atau modul adalah paket-paket yang umumnya digunakan dalam pemrograman untuk menyederhanakan implementasi kode. Dalam konteks pemodelan ini, pustaka yang dibutuhkan termasuk pandas, sklearn.feature, dan sklearn.metrics. Adapun bentuk dari model library dapat dilihat pada gambar 4.1

```
# import package
import nltk
import pandas as pd
import re
import numpy as np
import string
import os
import matplotlib
import seaborn as sns

# import module
from sklearn import preprocessing, model_selection, naive_bayes, svm
from sklearn.feature_extraction.text import CountVectorizer, TfidfVectorizer
from sklearn.metrics import accuracy_score
from matplotlib import pyplot as plt
from sklearn.metrics import confusion_matrix, accuracy_score, classification_report
```

Gambar 4. 1 Library Processing Classification

4.6.2 Teks Preprocessing

Tahapan teks preprocessing pada penelitian ini berfungsi untuk menghilangkan data mentah yang mengandung data yang tidak lengkap dan memiliki format yang tidak konsisten. Adapun beberapa tahapan data preprocessing dapat dilihat pada penjabaran berikut ini.

a. Cleaning

Pada fase awal pengolahan teks, akan dilakukan tindakan pembersihan data seperti menghapus tanda baca, angka, simbol, URL, dan username yang terdapat dalam teks. Adapun bentuk proses cleaning dapat dilihat pada gambar 4.2

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

```
#remove URL
def deURL(yt):
    yt = re.sub('http://\S+|https://\S+', '', yt)
    return yt

# remove URL
dataset['remove_url'] = dataset['tweet'].apply(lambda x: deURL(x))
# wajib

#remove hashtag
def deHashtag(yt):
    yt = re.sub(r'#[A-Za-z0-9]+', '', yt)
    return yt

# remove hashtag
dataset['remove_hashtag'] = dataset['remove_url'].apply(lambda x: deHashtag(x))
```

Gambar 4. 2 Tahap Cleaning

b. Case folding

Tahap proses teks selanjutnya akan melakukan perubahan teks menjadi huruf kecil. Adapun bentuk proses case folding dapat dilihat pada gambar 4.3

```
# uppercase to lowercase
def case_folding(yt):
    yt = yt.lower()
    return yt

# Case folding
dataset['casefolding'] = dataset['remove_tb'].apply(lambda x: case_folding(x))
```

Gambar 4. 3 Tahap Case Folding

c. Tokenizing

Tahap proses teks selanjutnya akan melakukan pemisahan teks kalimat menjadi berupa potongan kata- kata. Adapun bentuk proses tokenizing dapat dilihat pada gambar 4.4

```
# mengubah kalimat menjadi potongan kata-kata
def tokenize(yt):
    tokens = re.split('\W+', yt)
    return tokens

# Tokenizing
dataset['tokenisasi'] = dataset['casefolding'].apply(lambda x: tokenize(x))
```

Gambar 4. 4 Tahap Tokenizing

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

d. Normalisasi

Langkah selanjutnya dalam proses teks akan mengubah kata-kata yang bersifat tidak formal atau slang menjadi kata-kata yang lebih formal atau baku. Adapun bentuk proses normalisasi dapat dilihat pada gambar 4.5

```
# Normalisasi
kamus_normalization = pd.read_excel('/gdrive/MyDrive/Dataset TA/kamus normalization.xlsx')

def normalization(words):
    for index in range(0,len(words)):
        indx_kata = kamus_normalization.kata==words[index]
        formal = list(set(kamus_normalization[indx_kata].normalisasi))
        if len(formal)==1:
            words[index]=formal[0]
    return words

dataset['normalisasi'] = dataset['tokenisasi'].apply(normalization)
```

Gambar 4. 5 Tahap Normalisasi

e. Stopword Removal

Langkah selanjutnya dalam proses teks adalah menghilangkan kata-kata umum yang sering muncul dalam teks dan dianggap tidak relevan dalam analisis teks. Adapun bentuk proses stopwords removal dapat dilihat pada gambar 4.6

```
# menghapus daftar kata-kata yang tidak bermakna
pip install sastrawi
from Sastrawi.StopWordRemover.StopWordRemoverFactory import StopWordRemoverFactory, StopWordRemover, ArrayDictionary

# Ambil Stopword bawaan
stop_factory = StopWordRemoverFactory()
more_stopword = ["aq", "ada", "adalah", "adanya", "adapun", "agak", "agakny", "agar", "akan", "akankah", "akhir",
"akhiri", "akhirnya", "aku", "ak", "akulah", "amat", "amatlah", "anda", "andalah", "antar", "antara",
"antaranya", "apa", "apaan", "apabila", "apakah", "apalagi", "apatah", "arti", "artinya", "asal",
"asalkan", "atas", "atau", "ataukah", "ataupun", "awal", "awalnya", "b", "bagai", "bagaikan",
"bagaimana", "bagaimanakah", "bagaimanapun", "bagaimanakah", "bagi", "bagian", "bahkan", "bahwa",
"bahwesanya", "bahwesanya", "baik", "baiklah", "bakal", "bakalan", "baik", "banyak", "bapek",
"baru", "bawah", "beberapa", "begini", "beginian", "beginilah", "beginilah", "begitu", "begitukah",
"begitulah", "begitupun", "bekerja", "belakang", "belakangan", "belum", "belumlah", "benar",
"benarkah", "benarlah", "berada", "berakhir", "berakhirilah", "berakhirnya", "berapa", "berapakah",
"berapalah", "berapapun", "berarti", "berawal", "berbagai", "berdatangan", "beri", "berikan",
"berikut", "berikutnya", "berjumlah", "berkali-kali", "berkata", "berkehendak", "berkeinginan",
```

Gambar 4. 6 Tahap Stopword Removal

f. Stemming

Tahap proses teks yang terakhir ini akan menghilangkan imbuhan, sisipan dan akhiran dalam sebuah teks. Adapun bentuk proses stemming dapat dilihat pada gambar 4.7



Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

```
# Stemming
!pip install Sastrawi
from Sastrawi.Stemmer.StemmerFactory import StemmerFactory

factory = StemmerFactory()
stemmer = factory.create_stemmer()
ps = nltk.PorterStemmer()

def stemming(d):
    return [stemmer.stem(word) for word in d]
```

Gambar 4. 7 Tahap Stemming

Adapun Output dari teks preprocessing dapat dilihat pada gambar 4.8

tweet	remove_url	remove_hashtag	remove_emot	remove_angka	remove_at	remove_tb	casefolding	tokenisasi	normalisasi	stopwordremoval	stemming
Ulama dan Santri dukung Ganjar Capres 2024 https://t.co/YVVF7eVT	Ulama dan Santri dukung Ganjar Capres 2024	Ulama dan Santri dukung Ganjar Capres 2024	Ulama dan Santri dukung Ganjar Capres 2024	Ulama dan Santri dukung Ganjar Capres	Ulama dan Santri dukung Ganjar Capres	Ulama dan Santri dukung Ganjar Capres	ulama dan santri dukung ganjar capres	[ulama, dan, santri, dukung, ganjar, capres.]	[ulama, dan, santri, dukung, ganjar, capres.]	[ulama, santri, dukung, capres.]	[ulama, santri, dukung, capres.]
Hasil Survei Terbaru apres 2024: Ganjar Vs Anies Vs Prabowo - CNBC Indonesia https://t.co/AcMTg5gHs2 #Jokowinomics #BersamaPrabowo #MenataMasaDepan	Hasil Survei Terbaru Capres 2024: Ganjar Vs Anies Vs Prabowo - CNBC Indonesia	Hasil Survei Terbaru Capres 2024: Ganjar Vs Anies Vs Prabowo - CNBC Indonesia	Hasil Survei Terbaru Capres 2024: Ganjar Vs Anies Vs Prabowo - CNBC Indonesia	Hasil Survei Terbaru Capres 2024: Ganjar Vs Anies Vs Prabowo - CNBC Indonesia	Hasil Survei Terbaru Capres 2024: Ganjar Vs Anies Vs Prabowo - CNBC Indonesia	Hasil Survei Terbaru Capres 2024: Ganjar Vs Anies Vs Prabowo - CNBC Indonesia	hasil survei terbaru capres ganjar vs anies vs prabowo cncb indonesia	[hasil, survei, terbaru, capres, ganjar, vs, anies, vs, prabowo, cncb, indonesia.]	[hasil, survei, terbaru, capres, ganjar, vs, anies, vs, prabowo, cncb, indonesia.]	[hasil, survei, terbaru, capres, vs, anies, vs, prabowo, cncb, indonesia.]	[hasil, survei, terbaru, capres, vs, anies, vs, prabowo, cncb, indonesia]
@AUdinunyu27760 Bukan ketar ketir lagi sekarang jadi ngos ngosan Capres dan Cawapres lawannya berat...dulu di ajak bangun negara oleh padip damolnoi Ganiar	@AUdinunyu27760 Bukan ketar ketir lagi sekarang jadi ngos ngosan Capres dan Cawapres lawannya berat...dulu di ajak bangun negara oleh padip	@AUdinunyu27760 Bukan ketar ketir lagi sekarang jadi ngos ngosan Capres dan Cawapres lawannya berat...dulu di ajak bangun negara oleh padip	@AUdinunyu27760 Bukan ketar ketir lagi sekarang jadi ngos ngosan Capres dan Cawapres lawannya berat...dulu di ajak bangun negara oleh padip	@AUdinunyu27760 Bukan ketar ketir lagi sekarang jadi ngos ngosan Capres dan Cawapres lawannya berat...dulu di ajak bangun negara oleh padip	@AUdinunyu27760 Bukan ketar ketir lagi sekarang jadi ngos ngosan Capres dan Cawapres lawannya berat...dulu di ajak bangun negara oleh padip	Bukan ketar ketir lagi sekarang jadi ngos ngosan Capres dan Cawapres lawannya berat...dulu di ajak bangun negara oleh padip	bukan ketar ketir lagi sekarang jadi ngos ngosan capres dan cawapres lawannya berat dulu di ajak bangun negara oleh padip	[, bukan, ketar, ketir, lagi, sekarang, jadi, ngos, ngosan, capres, dan, cawapres, lawannya, berat, dulu, di, ajak, bangun, negara, oleh, padip, neaora, oleh.]	[, bukan, ketar, ketir, lagi, sekarang, jadi, ngos, ngosan, capres, dan, cawapres, lawannya, berat, dulu, di, ajak, bangun, negara, oleh, padip, neaora, oleh.]	[, ketar, ketir, ngos, ngos, capres, capres, lawanya, berat, ajak, bangun, negara, padip, damolnoi.]	[, ketar, ngos, capres, cawapres, lawa, berat, ajak, bangun neaora.]

Gambar 4. 8 Output Teks Preprocessing

4.6.3 Pembobotan TF-IDF

Langkah selanjutnya adalah membuat model TF-IDF untuk mewakili setiap tweet sebagai vektor fitur numerik. TF-IDF ini akan memberikan bobot pada kata-kata berdasarkan seberapa sering kata tersebut muncul dalam tweet tertentu dan seberapa unik kata tersebut dalam seluruh korpus tweet. Adapun bentuk dari pembobotan model TF-IDF dapat dilihat pada gambar 4.9

```
from sklearn.feature_extraction.text import TfidfVectorizer
import pandas as pd

## Skip langkah balancing kalau mau memproses data tanpa balancing (disable 2 baris di bawah ini)
x_train = Balanced['Steaming']
y_train = Balanced['polarity']

# Langkah tanpa balancing langsung saja ke bawah ini
corpus = x_train.tolist() # Memastikan bahwa corpus telah didefinisikan
corpus_combined = [" ".join(doc) for doc in corpus]
vectorizer = TfidfVectorizer()
X = vectorizer.fit_transform(corpus_combined)
feature_names = vectorizer.get_feature_names_out()
tfidf_df = pd.DataFrame(data=X.toarray(), columns=feature_names)
# print(tfidf_df)
```

Gambar 4. 9 Pembobotan TF-IDF

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

Adapun Output dari proses TF-IDF diatas dapat dilihat pada gambar 4.10

tfidf_df	aahuuuyyy	abadi	abdi	abdul	abdurrahman	abiiss	abu	acara	achet	acung	...	ykn	yogyakarta	youtube	yuk	yukk	yupz	yusril	zaman	zinedine	zionis	
0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	
1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	
2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	
3	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	
4	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	
...
1174	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	
1175	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	
1176	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	
1177	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	
1178	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	

Gambar 4. 10 Output TF-IDF

4.6.4 Pembagian Data

Data yang digunakan diperoleh dari Ilham dkk (2023), dengan statistik sebagai mana tabel 4.1 berikut.

Tabel 4.1. Statistik Data Sentimen Penyelenggara Piala Dunia U20

No	Deskripsi	Kelas Positif	Kelas Netral	Kelas Negatif
1.	Data Training	313	625	2146
2.	Data Testing	72	156	544
	Jumlah	385	781	2690

Data Training Pelatihan Model Naïve Bayes

Model klasifikasi Naive Bayes dilatih menggunakan data latih. Pada tahap ini Naive Bayes menghitung probabilitas kemunculan kata-kata dalam setiap kelas sentimen (negatif, positif, netral) berdasarkan data latih. Adapun bentuk model Naive Baye dapat dilihat pada gambar 4.12

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

```

from sklearn.naive_bayes import MultinomialNB
from sklearn.feature_extraction.text import TfidfVectorizer
import pandas as pd

# Pastikan X_val sudah di-transform dengan TfidfVectorizer
X_val_transformed = vectorizer.transform(X_val_combined)

# Inisialisasi dan pelatihan model Multinomial Naive Bayes
clasfc = MultinomialNB()
cl = clasfc.fit(X, y_train)
Y_pred = cl.predict(X_val_transformed)
    
```

Gambar 4. 11 Model Naïve Bayes

4.6.5 Prediksi Sentimen

Setelah melatih model, model dapat digunakan untuk memprediksi sentimen dari tweet pada data uji. Model ini akan menilai probabilitas tweet masuk ke dalam setiap kelas sentimen berdasarkan perhitungan probabilitas Naive Bayes. Adapun bentuk proses prediksinya dapat dilihat pada gambar 4.13

```

# TESTING KEPADA DATA FINAL
X_test_combined = [" ".join(doc) for doc in X_test]
X_test_transformed = vectorizer.transform(X_test_combined)

Y_pred_test = cl.predict(X_test_transformed)
confusion_matrix(y_test, Y_pred_test)
    
```

Gambar 4. 12 Prediksi Sentimen

Adapun output dari proses prediksi sentimen diatas dapat dilihat pada gambar 4.14

```

print(classification_report(y_test, Y_pred_test))
    
```

	precision	recall	f1-score	support
negative	0.90	0.67	0.77	57
netral	0.58	0.60	0.59	25
positive	0.73	0.88	0.80	67
accuracy			0.75	149
macro avg	0.74	0.72	0.72	149
weighted avg	0.77	0.75	0.75	149

Gambar 4. 13 Output Prediksi Sentimen

4.7 Pengujian

Agar dapat menilai performa model yang telah dibuat, diperlukan tahapan pengujian. Melalui proses pengujian ini, dapat diketahui sejauh mana kinerja model dan mendapatkan informasi mengenai tingkat akurasi yang berhasil

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

dicapai. Skema pengujian dilakukan untuk melihat performa klasifikasi dari masing-masing metode (Lexicon-based dan Naïve Bayes).

Untuk pengujian Lexicon-based, karena hanya menggunakan rule based berdasarkan kata-kata yang ada di dalam tweet berada di dalam kamus positif atau negatif saja, maka pengujian validasi dilakukan terhadap Data Training, yaitu 90% dari keseluruhan data. Sedangkan Data testing sebanyak 10%, selanjutnya akan digunakan untuk pengujian kedua metode.

Selanjutnya, untuk metode Naïve Bayes, pengujian dilakukan dengan skema berikut. Pada saat proses training metode Naïve Bayes (pembentukan matriks probabilitas fitur), data training (yang 90% pada eksperimen Lexicon-based), dipecah kembali dengan komposisi 90:10, untuk diambil 10% sebagai Data Validasi. Hal ini diperlukan untuk melakukan beberapa langkah optimasi untuk peningkatan performa klasifikasi.

Hasil pengujian yang dilakukan dituliskan di dalam Tabel 4.17 berikut. Data testing yang sama dipakai sebagai pengujian akhir dari masing-masing metode untuk dibandingkan performanya.

Tabel 4. 16 Eksperimen Motode Naïve Bayes dan Lexicon

Eksperimen	Hasil Pengujian			
	Data Validasi		Data Testing	
	F1-score	Accuracy	F1-score	Accuracy
Metode Lexicon	28,48%	49,32%	30,72%	54,53%
Metode NB Tanpa Balancing	29,24%	69,58%	29,81%	70,98%
Metode NB Dengan Balancing	81,48%	80,25%	61,74%	63,08%

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

4.7.1 Analisa Lexicon Based

Metode Lexicon-based digunakan untuk menentukan kelas pada dataset. Kelas yang digunakan pada penelitian ini adalah kelas positif, negatif dan netral. Hasil yang diperoleh dalam Tabel 4.16 menunjukkan metode lexicon based tidak akurat dalam memberikan hasil klasifikasi dataset terbukti dengan nilai F1-Score yang dihasilkan hanya mencapai angka 30,72 % dan hasil accuracy 54,53% pada pengujian data testing.

4.7.2 Data Balancing (penyeimbangan data)

Setup experiment dilakukan untuk menemukan model yang melakukan klasifikasi paling optimal terhadap data yang dilakukan pengujian. Pada penelitian ini metode yang diterapkan adalah metode *naive bayes* dan metode *lexicon based*, dimana jumlah data yang digunakan pada penelitian diberikan variasi yakni jumlah dataset yang seimbang (*balancing*) dengan jumlah dataset yang tidak seimbang (*unbalancing*).

Berdasarkan pengujian terhadap data testing yang dilakukan maka ditemukan hasil akhir pada experiment yaitu model lebih optimal dalam melakukan klasifikasi pada dataset yang seimbang (*balancing*) dengan nilai F1-Score 61,74% dan nilai accuracy 63,08% jika dibandingkan dengan dengan jumlah data yang tidak seimbang yang hanya mendapatkan nilai F1-Score 29,08% dan nilai accuracy 70,98%. Adapun hasil akhir yang ditemukan dengan menggunakan metode Lexicon Based adalah nilai F1-Score 30,72% dan nilai accuracy 70,98%.

Berdasarkan hasil akhir dari setup experiment terhadap 2 metode yang diterapkan, maka dapat diambil sebuah kesimpulan bahwa model yang lebih baik dalam melakukan klasifikasi sentimen adalah model Naive Bayes Classifier dengan menggunakan jumlah dataset yang seimbang (*balancing*). Dengan *balancing data*, model NB yang diperoleh dapat ditingkatkan dengan signifikan dibandingkan menggunakan data training yang tidak diseimbangkan.

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

4.7.3 Kesimpulan hasil pengujian

Berdasarkan dari pengujian yang sudah dilakukan, dapat disimpulkan hasil sebagai berikut :

1. Metode lexicon based pada pengujian kali ini tidak optimal dalam memprediksi dan mengklasifikasi kelas terbukti dengan nilai F1-Score yang dihasilkan hanya mencapai angka 30,72 % dan hasil accuracy 54,53%.
2. Pengujian dilakukan dengan rasio 90%: 10% (training : testing) dan untuk data validasi diambil 10% pada dataset. Hasil pengujian mendapatkan hasil accuracy 63,09 % dan nilai F1-Score 61,74%.
3. Pada penelitian ini dilakukan perbandingan antara jumlah dataset yang seimbang dan tidak seimbang dalam melakukan klasifikasi, adapun hasil yang ditemukan adalah model lebih baik melakukan klasifikasi menggunakan data yang seimbang dengan nilai F1-Score 61,74% jika dibandingkan dengan jumlah dataset yang tidak seimbang yang hanya mendapatkan nilai F1-Score 29,81%.
4. Pada penelitian ini dilakukan pengujian terhadap dua metode yaitu metode naïve bayes classifer dan metode lexicon based dalam melakukan klasifikasi sentimen adapun hasil akhir dari penelitian ini menunjukkan bahwa naïve bayes classifer lebih baik dalam melakukan klasifikasi dengan nilai F1-Score 61,74% jika dibandingkan dengan hasil klasifikasi menggunakan metode lexicon yang hanya mendapatkan hasil F1-Score 29,81%.