

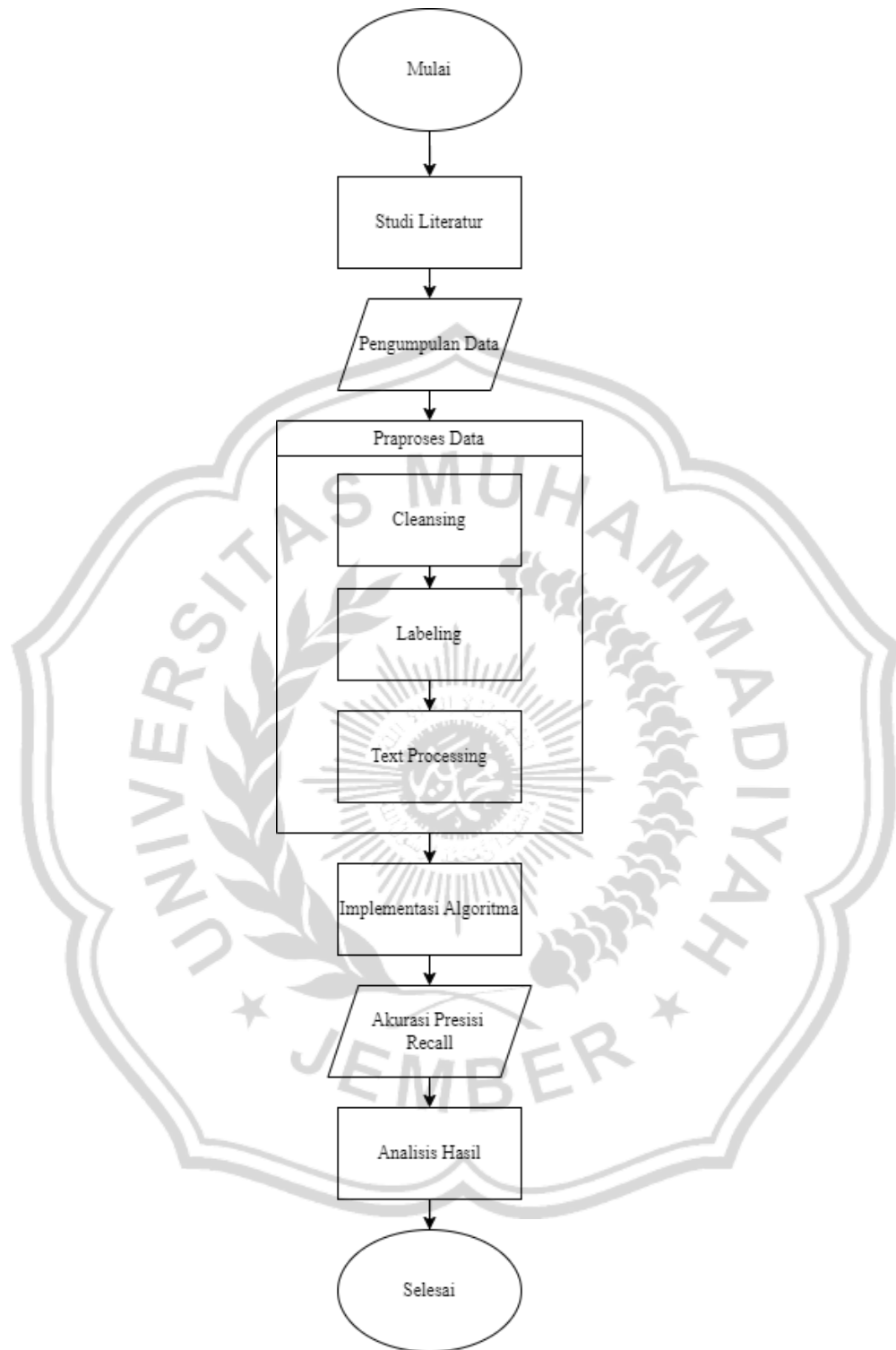
## BAB III METODE PENELITIAN

### 3.1 Tahapan Penelitian

Metode penelitian adalah prosedur atau pendekatan sistematis yang digunakan oleh peneliti untuk mengumpulkan, menganalisis, dan menyajikan data dalam rangka menjawab pertanyaan penelitian atau mencapai tujuan penelitian tertentu. Ada beberapa metode penelitian yang umum digunakan, seperti:

- a. Metode Kualitatif: Fokus pada pemahaman mendalam tentang fenomena sosial atau perilaku manusia. Pengumpulan data dilakukan melalui observasi, wawancara, atau analisis teks. Contohnya: studi kasus, etnografi, fenomenologi.
- b. Metode Kuantitatif: Berfokus pada pengukuran fenomena dengan menggunakan angka dan statistik untuk menganalisis data. Pengumpulan data dilakukan melalui survei, eksperimen, atau analisis data sekunder. Contohnya: survei, eksperimen kontrol acak, analisis regresi. Metode yang digunakan dalam penelitian ini adalah kuantitatif dan dilakukan secara bertahap dengan alur seperti pada gambar 3.1.

Dimulai dari studi literatur yang bertujuan untuk mendapat referensi dari penelitian terdahulu tentang cara melakukan sentimen analisis menggunakan metode *Naïve Bayes*. Dilanjutkan dengan pengumpulan data dengan metode *Scrapping* di sosial media *Twitter* yang nantinya akan mengalami serangkaian pra-proses agar terhindar dari data *noise* dan berubah bentuk sesuai kebutuhan model klasifikasi *Naïve Bayes*. Adapun hasil yang diharapkan adalah akurasi, presisi, dan *recall* dari model dan diakhiri dengan evaluasi model.



Gambar 3. 1 Tahapan Penelitian

### 3.2 Studi Literatur

Studi literatur merupakan tahapan penelitian yang melibatkan proses pencarian dan pengumpulan informasi dari berbagai literatur seperti buku, jurnal, tesis, dan lain-lain yang berkaitan dengan penelitian pengklasifikasian dan sentimen analisis yang menerapkan algoritma *Naïve Bayes*.

### 3.3 Pengumpulan Data

Dataset yang digunakan dalam penelitian ini adalah dataset sekunder yang diperoleh dari proses *scrapping* terhadap platform twitter berupa teks berjumlah 578 data dari bulan April sampai Juli 2023. Tabel berikut merupakan sampel data sebelum melalui tahap praproses.

Tabel 3. 1 Contoh Data Mentah

No.	Tweet
1	@Kanseulir Menyejukkan Luar Biasa Pembenci Anies Baswedan Pukulan Telak Atas Anies Baswedan Maju Di Pilpres
2	"Repost dari @partaiperindo Kebakaran terjadi di Gang Lapo Sona, Kelurahan Baringin, Kota Sibolga, Sumatera Utara. #PerindoUpdate #PartaiPerindo #Perindo #PerindoS16AP #PartaiPersatuanIndonesia #UntukIndonesiaSejahtera #Pemilu2024 #Pileg2024 #Pilkada2024 #Pilpres2024 <a href="https://t.co/zNHAIXLsFg">https://t.co/zNHAIXLsFg</a>
3	@72Gar7 @tvOneNews @Sigiwarmada Kasihan Kadrun yah partai nya sudah mau 10 tahun engga dapet jabatan Menteri, makanya kadrun sangat menderita....nanti 2024 kadrun kalah lagi Pilpres, menderita lagi 5 tahun, partai nya kadrun engga dapet jabatan Menteri.. □□□.. Hahaha
4	"Repost dari @partaiperindo Bantuan gerobak gratis dari Partai Perindo berdampak langsung meningkatkan perekonomian warga. #PerindoUpdate #PartaiPerindo #Perindo #PerindoS16AP #PartaiPersatuanIndonesia #UntukIndonesiaSejahtera #Pemilu2024 #Pileg2024 #Pilkada2024 #Pilpres2024 <a href="https://t.co/uUjFde6M5n">https://t.co/uUjFde6M5n</a>
5	wy Pemimpin Muda Inovatif dan Kreatif di dukung penuh generasi milenial Pemimpin muda yang punya banyak ide dan inovasi dalam memajukan daerah Suara semesta mendukung Cak Fauzi Calon Gubernur Jawa Timur di Pilpres
6	pingit imin setop bicara soal pilpres

### 3.4 Praproses Teks

Praproses teks perlu dilakukan untuk mempersiapkan dan menyelaraskan data teks mentah sebelum dilakukan analisis. Langkah-langkah seperti *case folding*, tokenisasi, penghapusan *stopwords*, *stemming*, dan pembobotan kata membantu mengurangi *noise* dan meningkatkan akurasi dalam mengekstraksi makna dan sentimen dari teks.

#### 3.4.1 Cleaning

Data yang akan digunakan adalah dataset dari kolom *Tweet*, atribut selain itu dianggap sebagai *noise* dan akan dibersihkan dengan cara dihapus dari tabel. Proses *Cleaning* juga menghapus tanda baca, *URL*, dan atribut lain yang tidak berkaitan dengan proses klasifikasi.

Tabel 3. 2 Contoh *Data Cleaning*

No.	Tweet
1	Menyejukkan Luar Biasa Pembenci Anies Baswedan Pukulan Telak Atas Anies Baswedan Maju Di Pilpres
2	Repost dari Kebakaran terjadi di Gang Lapo Sona Kelurahan Baringin Kota Sibolga Sumatera Utara
3	Kasihannya Kadrun yah partai nya sudah mau tahun engga dapet jabatan Menteri makanya kadrun sangat menderitananti kadrun kalah lagi
4	Repost dari Bantuan gerobak gratis dari Partai Perindo berdampak langsung meningkatkan perekonomian warga
5	Pemimpin Muda Inovatif dan Kreatif di dukung penuh generasi milenial Pemimpin muda yang punya banyak ide dan inovasi dalam memajukan daerah Suara semesta mendukung Cak Fauzi Calon Gubernur Jawa Timur di Pilpres
6	pingit imin setop bicara soal pilpres

#### 3.4.2 Labeling

Pada tabel 3.3 Proses pelabelan dilakukan secara manual ke dalam kategori positif (1) dan negatif (0). Untuk data keseluruhan dilabeli oleh Tia Dwi Putri Hari Nanda, S.Pd. selaku guru Bahasa Indonesia di SMPN 02 Botolinggo Bondowoso. Berikut sampel data yang sudah dilabeli:

Tabel 3. 3 Contoh Sampel Hasil Pelabelan.

Tweet	Label
Menyejukkan Luar Biasa Pembenci Anies Baswedan Pukulan Telak Atas Anies Baswedan Maju Di Pilpres	0
Repost dari Kebakaran terjadi di Gang Lapo Sona Kelurahan Baringin Kota Sibolga Sumatera Utara	0
Kasihannya Kadrun yah partai nya sudah mau tahun engga dapet jabatan Menteri makanya kadrun sangat menderitananti kadrun kalah lagi	0
Repost dari Bantuan gerobak gratis dari Partai Perindo berdampak langsung meningkatkan perekonomian warga	1
Pemimpin Muda Inovatif dan Kreatif di dukung penuh generasi milenial Pemimpin muda yang punya banyak ide dan inovasi dalam memajukan daerah Suara semesta mendukung Cak Fauzi Calon Gubernur Jawa Timur di Pilpres	1
pingit imin setop bicara soal pilpres	0

### 3.4.3 Text Preprocessing

Tahap *text preprocessing* melibatkan beberapa langkah untuk mengolah teks sebelum dilakukan klasifikasi agar bisa dikenali dan dibaca oleh komputer. Adapun langkah-langkah *text preprocessing* adalah sebagai berikut:

- Menurut (Gunawan et al., 2018) *case folding* merupakan proses mengubah semua huruf dalam suatu dokumen atau kalimat menjadi huruf kecil.

Tabel 3. 4 Contoh Hasil *Case folding*

Tweet	Label	Case Folding
Menyejukkan Luar Biasa Pembenci Anies Baswedan Pukulan Telak Atas Anies Baswedan Maju Di Pilpres	0	menyejukkan luar biasa tapi bagi pembenci anies baswedan ini adalah pukulan telak atas mereka yg ingin menjegal anies baswedan maju di pilpres
Repost dari Kebakaran terjadi di Gang Lapo Sona Kelurahan Baringin Kota Sibolga Sumatera Utara	0	repost dari kebakaran terjadi di gang lapo sona kelurahan baringin kota sibolga sumatera utara
Kasihannya Kadrun yah partai nya sudah mau tahun engga dapet jabatan Menteri makanya kadrun sangat menderitananti kadrun kalah lagi	0	kasihan kadrun yah partai nya sudah mau tahun engga dapet jabatan menteri makanya kadrun sangat menderitananti kadrun kalah lagi

Tweet	Label	Case Folding
Repost dari Bantuan gerobak gratis dari Partai Perindo berdampak langsung meningkatkan perekonomian warga	1	repost dari bantuan gerobak gratis dari partai perindo berdampak langsung meningkatkan perekonomian warga
Pemimpin Muda Inovatif dan Kreatif di dukung penuh generasi milenial Pemimpin muda yang punya banyak ide dan inovasi dalam memajukan daerah Suara semesta mendukung Cak Fauzi Calon Gubernur Jawa Timur di Pilpres	1	pemimpin muda inovatif dan kreatif di dukung penuh generasi milenial pemimpin muda yang punya banyak ide dan inovasi dalam memajukan daerah suara semesta mendukung cak fauzi calon gubernur jawa timur di pilpres
pingit cak imin untuk setop bicara soal pilpres	0	pingit cak imin untuk setop bicara soal pilpres

### b. *Tokenizing*

Tokenisasi adalah proses dalam pra-pemrosesan teks yang melibatkan pemecahan teks menjadi unit-unit yang lebih kecil yang disebut token. Token dapat berupa kata, kalimat, atau bahkan unit yang lebih kecil seperti karakter atau subkata, tergantung pada teknik tokenisasi yang digunakan.

*Tabel 3.5 Contoh Hasil Tokenizing*

Case Folding	Label	Token
menyejukkan luar biasa pembenci anies baswedan pukulan telak atas anies baswedan maju di pilpres	0	['menyejukkan', 'luar', 'biasa', 'pembenci', 'anies', 'baswedan', 'pukulan', 'telak', 'atas', 'anies', 'baswedan', 'maju', 'di', 'pilpres']
repost dari kebakaran terjadi di gang lapo sona kelurahan baringin kota sibolga sumatera utara	0	['repost', 'dari', 'kebakaran', 'terjadi', 'di', 'gang', 'lapo', 'sona', 'kelurahan', 'baringin', 'kota', 'sibolga', 'sumatera', 'utara']
kasihan kadrun yah partai nya sudah mau tahun engga dapet jabatan menteri	0	['kasihan', 'kadrun', 'yah', 'partai', 'nya', 'sudah', 'mau', 'tahun', 'engga', 'dapet', 'jabatan', 'menteri', 'makanya', 'kadrun',

makanya kadrun sangat menderitananti kadrun kalah lagi		'sangat','menderitananti','kadrun', 'kalah','lagi"pilpres', 'menderita', 'lagi']
repost dari bantuan gerobak gratis dari partai perindo berdampak langsung meningkatkan perekonomian warga	1	['repost', 'dari', 'bantuan', 'gerobak', 'gratis', 'dari', 'partai', 'perindo', 'berdampak"langsung','meningkatkan', 'perekonomian', 'warga']
Simpatisan Jokowi di Ponorogo mendeklarasikan dukungan untuk Prabowo Subianto pada Pilpres Gerindra yakin nanti Prabowo menang di Ponorogo	1	['simpatisan', 'jokowi', 'di', 'ponorogo', 'mendeklarasikan', 'dukungan', 'untuk', 'prabowo', 'subianto', 'pada', 'pilpres', 'gerindra', 'yakin', 'nanti', 'prabowo', 'menang', 'di', 'ponorogo']
pingit cak imin untuk setop bicara soal pilpres	0	['pingit', 'cak', 'imin', 'untuk', 'setop', 'bicara', 'soal', 'pilpres']

### c. Stopword Removal

Stopword removal adalah proses penghapusan kata-kata pengisi atau kata-kata umum yang sering muncul dalam teks tetapi tidak memiliki makna penting untuk pemrosesan teks atau analisis sentimen. Kata-kata ini disebut "stopwords" dan biasanya termasuk kata-kata seperti "dan", "atau", "dari", "yang", "saya", "kamu", dan lain sebagainya.

*Tabel 3. 6 Contoh Hasil Stopword Removal*

Token	Label	Stopword Removal
['menyejukkan', 'luar', 'biasa', 'pembenci', 'anies', 'baswedan', 'pukulan', 'telak' atas', 'anies', 'baswedan' maju', 'di', 'pilpres']	0	['menyejukkan', 'luar', 'biasa', 'pembenci', 'anies', 'baswedan', 'pukulan', 'telak' atas', 'anies', 'baswedan' maju', 'pilpres']
['repost', 'dari', 'kebakaran', 'terjadi', 'di', 'gang', 'lapo', 'sona', 'kelurahan', 'baringin', 'kota', 'sibolga', 'sumatera', 'utara']	0	['repost', 'dari', 'kebakaran', 'terjadi', 'gang', 'lapo', 'sona', 'kelurahan', 'baringin', 'kota', 'sibolga', 'sumatera', 'utara']
['kasihan', 'kadrun', 'yah', 'partai', 'nya', 'sudah', 'mau', 'tahun', 'engga', 'dapet', 'jabatan', 'menteri', 'makanya', 'kadrun', 'sangat', 'menderitananti', 'kadrun', 'kalah', 'lagi']	0	['kasihan', 'kadrun', 'yah', 'partai', 'nya', 'sudah', 'mau', 'tahun', 'engga', 'dapet', 'jabatan', 'menteri', 'makanya', 'kadrun', 'sangat', 'menderitananti', 'kadrun', 'kalah', 'lagi']
['repost', 'dari', 'bantuan', 'gerobak', 'gratis', 'dari', 'partai', 'perindo',	1	['repost', 'dari', 'bantuan', 'gerobak', 'gratis', 'dari', 'partai', 'perindo',

'berdampak', 'meningkatkan', 'warga']	'langsung', 'perekonomian', 'warga']		'berdampak', 'meningkatkan', 'warga']
['simpatisan', 'ponorogo', 'dukungan', 'subianto', 'gerindra', 'prabowo', 'ponorogo']	'jokowi', 'di', 'mendeklarasikan', 'untuk', 'prabowo', 'pada', 'pilpres', 'yakin', 'nanti', 'menang', 'di', 'ponorogo']	1	['simpatisan', 'jokowi', 'ponorogo', 'mendeklarasikan', 'dukungan', 'untuk', 'prabowo', 'subianto', 'pilpres', 'gerindra', 'yakin', 'nanti', 'prabowo', 'menang', 'ponorogo']
['pingit', 'setop', 'bicara', 'soal', 'pilpres']	'cak', 'imin', 'untuk', 'setop', 'bicara', 'soal', 'pilpres']	0	['pingit', 'cak', 'imin', 'untuk', 'setop', 'bicara', 'soal', 'pilpres']

Tujuan dari penghapusan stopwords adalah untuk meningkatkan kualitas pemrosesan teks atau analisis sentimen dengan fokus pada kata-kata yang lebih penting dan memiliki makna yang lebih signifikan. Dengan menghapus stopwords, kita dapat menghilangkan derau dan mengoptimalkan analisis teks dengan mempertimbangkan hanya kata-kata yang lebih informatif.

Tabel 3. 7 Contoh Daftar Kata Hubung yang Dihilangkan

Kata Hubung Yang Dihilangkan
'yang', 'untuk', 'pada', 'ke', 'para', 'namun', 'menurut', 'antara', 'dia', 'dua', 'ia', 'seperti', 'jika', 'jika', 'sehingga', 'kembali', 'dan', 'tidak', 'ini', 'karena', 'kepada', 'oleh', 'saat', 'harus', 'sementara', 'setelah', 'belum', 'kami', 'sekitar', 'bagi', 'serta', 'di', 'dari', 'telah', 'sebagai', 'masih', 'hal', 'ketika', 'adalah', 'itu', 'dalam', 'bisa', 'bahwa', 'atau', 'hanya', 'kita', 'dengan', 'akan', 'juga', 'ada', 'mereka', 'sudah', 'saya', 'terhadap', 'secara', 'agar', 'lain', 'anda', 'begitu', 'mengapa', 'kenapa', 'yaitu', 'yakni', 'daripada', 'itulah', 'lagi', 'maka', 'tentang', 'demi', 'di mana', 'kemana', 'pula', 'sambil', 'sebelum', 'sesudah', 'supaya', 'guna', 'kah', 'pun', 'sampai', 'sedangkan', 'selagi', 'sementara', 'tetapi', 'apakah', 'kecuali', 'sebab', 'selain', 'seolah', 'seraya', 'seterusnya', 'tanpa', 'agak', 'boleh', 'dapat', 'dsb', 'dst', 'dll', 'dahulu', 'dulunya', 'anu', 'demikian', 'tapi', 'ingin', 'juga', 'nggak', 'mari', 'nanti', 'melainkan', 'oh', 'ok', 'seharusnya', 'sebetulnya', 'setiap', 'setidaknya', 'sesuatu', 'pasti', 'saja', 'toh', 'ya', 'walau', 'tolong', 'tentu', 'amat', 'apalagi', 'bagaimanapun'

#### d. *Stemming*

Stemming adalah proses pemotongan atau penghapusan awalan atau akhiran kata dalam bahasa agar hanya menyisakan bentuk dasar atau kata dasar (stem) dari kata tersebut.



Tabel 3. 8 Contoh Daftar Kalimat yang sudah Disesuaikan

Stopword Removal	Label	Stemming
['menyejukkan', 'luar', 'biasa', 'pembenci', 'anies', 'baswedan', 'pukulan', 'telak' 'atas', 'anies', 'baswedan' 'maju', 'pilpres']	0	sejuk luar biasa benci anies baswedan pukul telak anies baswedan maju pilpres
['repost', 'dari', 'kebakaran', 'terjadi', 'gang', 'lapo', 'sona', 'kelurahan', 'baringin', 'kota', 'sibolga', 'sumatera', 'utara']	0	repost dari bakar jadi gang lapo sona lurah baringin kota sibolga sumatera utara
['kasihan', 'kadrun', 'yah', 'partai', 'nya', 'sudah', 'mau', 'tahun', 'engga', 'dapat', 'jabatan', 'menteri', 'makanya', 'kadrun', 'sangat', 'menderitananti', 'kadrun', 'kalah', 'lagi']	0	kasihan kadrun yah partai sudah mau tahun engga dapat jabat menteri makanya kadrun sangat menderitananti kadrun kalah lagi
['repost', 'dari', 'bantuan', 'gerobak', 'gratis', 'dari', 'partai', 'perindo', 'berdampak', 'langsung', 'meningkatkan', 'perekonomian', 'warga']	1	repost dari bantu gerobak gratis dari partai perindo dampak langsung tingkat ekonomi warga
['simpatisan', 'jokowi', 'ponorogo', 'mendeklarasikan', 'dukungan', 'untuk', 'prabowo', 'subianto', 'pilpres', 'gerindra', 'yakin', 'nanti', 'prabowo', 'menang', 'ponorogo']	1	simpatisan jokowi ponorogo deklarasi dukung untuk prabowo subianto pilpres gerindra yakin nanti prabowo menang ponorogo
['pingit', 'cak', 'imin', 'untuk', 'setop', 'bicara', 'soal', 'pilpres']	0	pingit cak imin untuk setop bicara soal pilpres

### 3.5 TF-IDF

(Robertson, 2004) menuturkan bahwa prinsip *TF-IDF* adalah penentuan frekuensi relatif suatu kata kemudian dibandingkan dengan proporsi kata tersebut pada seluruh dokumen. Pada penelitian ini, perhitungan *TF-IDF* dilakukan menggunakan *library sklearn.feature\_extraction* pada python. Berikut contoh perhitungan *TF* menurut (Ruli Siregar et al., 2017).

$$tf_{t,d} = \frac{n_{t,d}}{N} \dots\dots\dots (3.1)$$

dimana:

$n_{t,d}$  = nilai kata yang muncul

$N$  = total kata pada dokumen

$tf_{t,d}$  = frekuensi kemunculan kata pada satu dokumen

Selanjutnya menghitung *IDF* untuk mengukur pentingnya sebuah kata. Siregar menurukan perhitungan *IDF* sebagai berikut:

$$idf_d = \log \left( \frac{N}{df} \right) \dots \dots \dots (3.2)$$

Keterangan:

$N$  = total dokumen

$df$  = banyak dokumen yang mengandung term tersebut

$idf_d$  = jumlah kemunculan suatu kata dalam dokumen

Berikut adalah contoh perhitungan bobot *TF-IDF*:

Tabel 3. 9 Daftar Dokumen

Dokumen	Teks	Sentimen
D1	sejuk luar biasa benci anies baswedan pukul telak anies baswedan maju pilpres	Negatif
D2	repost dari bakar jadi gang lapo sona lurah baringin kota sibolga sumatera utara	Negatif
D3	kasihan kadrun yah partai sudah mau tahun engga dapet jabat menteri makanya kadrun sangat menderitananti kadrun kalah lagi	Negatif
D4	repost dari bantu gerobak gratis dari partai perindo dampak langsung tingkat ekonomi warga	Positif
D5	simpatisan jokowi di ponorogo deklarasi dukung untuk prabowo subianto pada pilpres gerindra yakin nanti prabowo menang di ponorogo	Positif
D6	pingit cak imin untuk setop bicara soal pilpres	negatif

Keterangan :

D1: Dokumen 1

D2: Dokumen 2

D3: Dokumen 3

D4: Dokumen 4

D5: Dokumen 5

D6: Dokumen 6

Langkah pertama adalah mencari nilai *TF* (*Term Frequency*) pada kata 'sejuk' menggunakan rumus berikut:

$$tf_{t,d} = \frac{n_{t,d}}{N} = \frac{1}{12} = 0.083 \dots \dots \dots (3.3)$$

Tabel 3. 10 Contoh Perhitungan Term Frequency (TF)

Term	Dokumen				TF			
	D1	D2	D3	D4	TF 1	TF 2	TF 3	TF 4
Sejuk	1	0	0	0	0,0833	0	0	0
Luar	1	0	0	0	0,0833	0	0	0
Biasa	1	0	0	0	0,0833	0	0	0
Benci	1	0	0	0	0,0833	0	0	0
Anies	2	0	0	0	0,1667	0	0	0
baswedan	2	0	0	0	0,1667	0	0	0
Pukul	1	0	0	0	0,0833	0	0	0
Telak	1	0	0	0	0,0833	0	0	0
Maju	1	0	0	0	0,0833	0	0	0
pilpres	1	0	0	0	0,0833	0	0	0

Langkah selanjutnya adalah menghitung nilai *IDF* (*Inverse Document Frequency*) untuk kata 'sejuk' sebagai berikut:

$$IDF = \log \frac{N}{df} \dots \dots \dots (3.4)$$

N = Total Dokumen (6)

Df = jumlah kemunculan kata "sejuk" pada seluruh dokumen (1).

$$\frac{N}{Df} = \frac{6}{1} = 6$$

Log (6) = 0,77

Tabel 3. 11 Contoh Perhitungan Inverse Document Frequency (IDF)

<i>Term</i>	<i>DF</i>	<i>N/DF</i>	<i>IDF</i>
Sejuk	1	6	0,77
Luar	1	6	0,77
Biasa	1	6	0,77
Benci	1	6	0,77
Anies	2	3	0,77
Baswedan	2	3	0,77
Pukul	1	6	0,77
Telak	1	6	0,77
Maju	1	6	0,77
Pilpres	1	6	0,77

Lalu, langkah terakhir adalah mengalikan nilai *TF* dengan *IDF*. Berikut adalah menghitung nilai *TF-IDF* pada kata “sejuk” pada dokumen 1:

$$0,083 \times 0,77 = 0,06391$$

Tabel 3. 12 Daftar Nilai TF-IDF

<i>Term</i>	<i>TF-IDF</i>			
	<b>D1</b>	<b>D2</b>	<b>D3</b>	<b>D4</b>
Sejuk	0,06391	0,00	0,00	0,00
Luar	0,06391	0,00	0,00	0,00
Biasa	0,06391	0,00	0,00	0,00
Benci	0,06391	0,00	0,00	0,00
Anies	0,06391	0,00	0,00	0,00
baswedan	0,06391	0,00	0,00	0,00
Pukul	0,06391	0,00	0,00	0,00
Telak	0,06391	0,00	0,00	0,00
Maju	0,06391	0,00	0,00	0,00
pilpres	0,06391	0,00	0,00	0,00

### 3.6 Penerapan Metode

Langkah pertama adalah memuat dataset yang sudah dipreproses. Dataset ini harus berisi dua kolom, yaitu kolom teks yang sudah dipreproses (misalnya ‘stemming’) dan kolom label klasifikasi yang sesuai (misalnya 'label'). Setelah itu,

dataset dibagi menjadi dua bagian, yaitu data pelatihan dan data pengujian. Data pelatihan digunakan untuk melatih model klasifikasi, sedangkan data pengujian digunakan untuk melakukan prediksi dan menghitung akurasi. Selanjutnya, dilakukan representasi numerik dari teks menggunakan metode *TF-IDF*. Hal ini dilakukan dengan menghitung nilai *TF-IDF* untuk setiap kata dalam dataset. Representasi ini akan digunakan sebagai fitur untuk melatih model klasifikasi. Setelah itu, model klasifikasi Naive Bayes dibuat dan dilatih menggunakan data pelatihan dan representasi numerik TF-IDF. Model ini akan belajar pola dari data pelatihan untuk melakukan klasifikasi. Berikut rumus *Naïve Bayes*:

$$P(c|X) = \frac{P(x|C)P(c)}{P(x)} \dots\dots\dots (3.5)$$

Keterangan :

$P(c|X)$  : Peluang terjadinya  $c$  berdasarkan kondisi  $X$  (posteriori prob)

$P(c)$  : Peluang terjadinya  $c$  (prior prob)

$P(X|C)$  : Peluang terjadinya  $X$  berdasarkan kondisi pada hipotesis  $C$

$P(X)$  : Peluang terjadinya  $X$

Kemudian, dilakukan prediksi menggunakan model yang sudah dilatih pada data pengujian. Model akan memberikan prediksi label klasifikasi untuk setiap data pengujian.

1. Hitung Probabilitas Kemunculan Kata dalam Kelas Sentimen (Menggunakan nilai TF-IDF)

$$P(\text{sejuk}|\text{positif}) = 0,06391$$

$$P(\text{sejuk}|\text{negatif}) = 0$$

2. Menghitung probabilitas kelas:

$$P(\text{positif}) = \frac{\text{total dokumen positif}}{\text{total dokumen}} = \frac{2}{6} = 0,333 \dots\dots\dots (3.6)$$

$$P(\text{negatif}) = \frac{\text{total dokumen negatif}}{\text{total dokumen}} = \frac{2}{6} = 0,666 \dots\dots\dots (3.7)$$

3. Hitung probabilitas sentimen untuk data baru (dokumen 1)

$P(\text{positif} | \text{'sejuk'}, \text{'luar'}, \text{'biasa'}, \text{'anies'}, \text{'baswedan'}, \text{'pukul'}, \text{'telak'},$   
 $\text{'atas'}, \text{'jegal'}, \text{'anies'}, \text{'baswedan'}, \text{'maju'}, \text{'pilpres'})$

$$\begin{aligned} \text{Kelas Positif} &= P(\text{'sejuk'} | \text{Positif}) \times (\text{'luar'} | \text{Positif}) \\ &\times (\text{'biasa'} | \text{Positif}) \times (\text{'anies'} | \text{Positif}) \\ &\times (\text{'baswedan'} | \text{Positif}) \times (\text{'pukul'} | \text{Positif}) \\ &\times (\text{'telak'} | \text{Positif}) \times (\text{'atas'} | \text{Positif}) \\ &\times (\text{'jegal'} | \text{Positif}) \times (\text{'anies'} | \text{Positif}) \\ &\times (\text{'baswedan'} | \text{Positif}) \times (\text{'maju'} | \text{Positif}) \\ &\times (\text{'pilpres'} | \text{Positif}) \end{aligned}$$

$$\text{Kelas Positif} = \frac{0}{0,333} = 0 \dots\dots\dots (3.8)$$

$$\begin{aligned} \text{Kelas Negatif} &= P(\text{'sejuk'} | \text{Negatif}) \times (\text{'luar'} | \text{Negatif}) \\ &\times (\text{'biasa'} | \text{Negatif}) \times (\text{'anies'} | \text{Negatif}) \\ &\times (\text{'baswedan'} | \text{Negatif}) \times (\text{'pukul'} | \text{Negatif}) \\ &\times (\text{'telak'} | \text{Negatif}) \times (\text{'atas'} | \text{Negatif}) \\ &\times (\text{'jegal'} | \text{Negatif}) \times (\text{'anies'} | \text{Negatif}) \\ &\times (\text{'baswedan'} | \text{Negatif}) \times (\text{'maju'} | \text{Negatif}) \\ &\times (\text{'pilpres'} | \text{Negatif}) \end{aligned}$$

$$\begin{aligned} \text{Kelas Negatif} &= 0,06391 \times 0,06391 \times 0,06391 \times 0,06391 \times \\ &0,06391 \times 0,06391 \times 0,06391 \times 0,06391 \times 0,06391 \times 0,06391 \times \\ &0,06391 \times 0,06391 = \frac{4,64E-15}{0,666} = 6,97E - 15 \dots\dots\dots (3.9) \end{aligned}$$

4. Hitung probabilitas sentimen untuk dokumen 4:

$P(\text{positif} | \text{'repost'}, \text{'dari'}, \text{'bantu'}, \text{'gerobak'}, \text{'gratis'}, \text{'dari'}, \text{'partai'},$   
 $\text{'perindo'}, \text{'dampak'}, \text{'langsung'}, \text{'tingkat'}, \text{'ekonomi'}, \text{'warga'})$

$$\begin{aligned} \text{Kelas Positif} &= P(\text{'bantu'} | \text{Positif}) \times (\text{'gerobak'} | \text{Positif}) \times \\ &\times (\text{'partai'} | \text{Positif}) \times (\text{'perindo'} | \text{Positif}) \\ &\times (\text{'dampak'} | \text{Positif}) \times (\text{'langsung'} | \text{Positif}) \\ &\times (\text{'tingkat'} | \text{Positif}) \times (\text{'ekonomi'} | \text{Positif}) \\ &\times (\text{'warga'} | \text{Positif}) \end{aligned}$$

$$\begin{aligned} \text{Kelas Positif} &= 0,0502 \times 0,0502 \times 0,0502 \times 0,0502 \times 0,0502 \times \\ &0,0502 \times 0,0502 \times 0,0502 \times 0,0502 = \frac{2,01431E-12}{0,666} = 3,02E - 12 \dots\dots (3.10) \end{aligned}$$

$$\begin{aligned}
 \text{Kelas Negatif} &= P('bantu'|Negatif) \times ('gerobak'|Negatif) \times \\
 &\quad \times ('partai'|Negatif) \times ('perindo'|Negatif) \\
 &\quad \times ('dampak'|Negatif) \times ('langsung'|Negatif) \\
 &\quad \times ('tingkat'|Negatif) \times ('ekonomi'|Negatif) \\
 &\quad \times ('warga'|Negatif) \\
 \text{Kelas Negatif} &= \frac{0}{0,666} = 0 \dots\dots\dots (3.11)
 \end{aligned}$$

Tabel 3. 13 Hasil Klasifikasi

Dokumen	Negatif	Positif	Aktual	Klasifikasi	Prediksi
1	9,29E – 15	0	negatif	negatif	TN ( <i>True Negative</i> )
2	6,23763E-19	0	negatif	negatif	TN ( <i>True Negative</i> )
3	0,002148007	0	negatif	negatif	TN ( <i>True Negative</i> )
4	0	3,02E-12	positif	positif	TP ( <i>True Positive</i> )
5	0,5123154	0	positif	negatif	FP ( <i>False Positive</i> )
6	0	0,1252333	negatif	positif	FN ( <i>False Negative</i> )

### 3.7 Evaluasi Hasil

Tahap terakhir adalah analisis performa dari klasifikasi yang dihitung dengan membandingkan prediksi model dengan label yang sebenarnya pada data pengujian. Akurasi merupakan persentase dari prediksi yang benar terhadap total data pengujian. Semakin tinggi nilai akurasi, semakin baik performa model dalam melakukan klasifikasi. Selain itu, ada unit pengukuran lain untuk mengevaluasi performa dari klasifikasi yaitu recall dan presisi. *Confusion Matrix* digunakan untuk menghitung ketiganya.

Tabel 3. 14 *Confusion Matrix 2x2*

		Kelas Prediksi	
		1	2
Kelas Aktual	1	TP	FP
	2	FN	TN

Keterangan :

- *True Positive* (TP) adalah jumlah dari kelas positif, yaitu kelas yang mempunyai jumlah instance lebih sedikit yang benar diklasifikasikan.
- *False Negative* (FN) adalah jumlah kelas positif yang salah diklasifikasikan dalam kelas negatif.
- *False Positive* (FP) adalah jumlah kelas negatif yang diklasifikasikan ke dalam kelas positif.
- *True Negative* (TN) adalah jumlah kelas negatif yang benar diklasifikasikan.

Berikut contoh *confusion matrix*:

Tabel 3.15 Contoh *Confusion Matrix*

Kelas Prediksi	Kelas Aktual	
	Positif	Negatif
Positif	TP = 72	FP = 9
Negatif	FN = 17	TN = 18

Berikut rumus dari Akurasi, Presisi dan Recall:

$$Akurasi = \frac{(TP + TN)}{TP + FN + FP + TN} = \frac{(72 + 18)}{72 + 17 + 9 + 18} \times 100\% = 77\%$$

$$Presisi = \frac{(TP)}{(TP + FP)} = \frac{(72)}{72 + 9} \times 100\% = 88\%$$

$$Recall = \frac{(TP)}{(TP + FN)} = \frac{(72)}{72 + 17} \times 100\% = 80\%$$