

### BAB III HASIL ANALISIS DAN PEMBAHASAN

#### 3.1 Pengumpulan Data

Proses pengumpulan data untuk penelitian ini menghasilkan dataset yang komprehensif terkait diskusi *Quick Count* Pemilihan Presiden Indonesia 2024 di platform X Twitter. Menggunakan *Tweet Harvest*, berhasil dikumpulkan total 2113 *tweet* yang relevan dalam rentang waktu 14-15 Februari 2024. Proses *crawling* data berlangsung selama 29 menit pada 21 April 2024, dari pukul 21:11 WITA hingga 21:40 WITA. Dari sampel data yang ditampilkan pada Tabel 3.1, terlihat keragaman sudut pandang dan topik diskusi seputar *Quick Count*, termasuk pembahasan hasil sementara dan pertanyaan tentang pengaruh Quick Count terhadap proses pemilu selanjutnya.

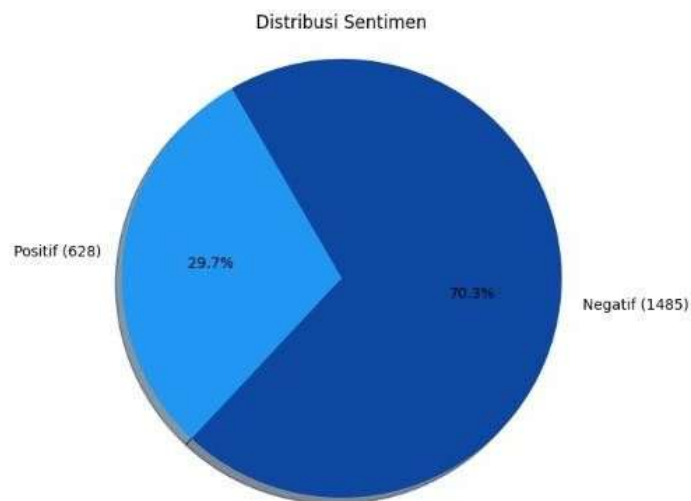
**Tabel 3.1** Hasil Pengumpulan Data

No	created_at	favorite_count	full_text	...	username
1	Wed Feb 14 22:24:25 +0000 2024	0	@m4retha Bilas muka gosok gigi liat quick count lagi	...	irfanalawiii
2	Wed Feb 14 22:24:19 +0000 2024	0	@nvtamarsaoly @convomfs Bener bgt. Ga jauh2 hasil real count nanti sama quick count. Udah nder capee ngomong sama yang ga mau kalah sipaling dicurangi dan tersakiti .	...	jisookimjichuu
3	Wed Feb 14 22:24:10 +0000 2024	1	Yg udh liat dirty vote masa percaya ma Quick Count .. Ayokk semangat para pejuang amin Real count masih berjalan	...	MantanKamu1993
...	...	...	...	...	...
2111	Wed Feb 14 16:29:27 +0000 2024	0	@andre_rosiade @prabowo @gibran_tweet Quick count hasil resmi kah? Lgsg lantik aja besok abis subuh	...	alexandersommy4
2112	Wed Feb 14 16:29:26 +0000 2024	1	@tanyarlfs Perasaan itu bukan pidato kemenangan deh tadi gw nonton fyi aja itu tuh emang udah direncanain kali acaranya buat ngelihat hasil quick count bareng timsesnya plis deh jangan kemakan judul berita minimal tonton sampe habis	...	ourrbee
2113	Wed Feb 14 16:29:25 +0000 2024	0	@m4retha Update 23.29: 02 unggul dengan suara diatas 50% pada 29 provinsi. Total quickcount di 38 provinsi + Luar Negeri: 01 Amin 32 50% 02 Pagi 51 21% 03 Gama 16 29%	...	msmrleia

Data yang terkumpul sebanyak 2113 *tweet* dengan berbagai variasi konten yang berhubungan dengan *Quick Count*, di mana hanya bagian *full\_text* yang diambil dari setiap *tweet*. Tahap selanjutnya data tersebut akan melalui proses *labeling* oleh seorang ahli pelabelan data yang memiliki kualifikasi dan pengalaman relevan.

### 3.2 Labeling Data

Dalam proses *labeling* data untuk klasifikasi teks *tweet*, *labeling* dilakukan oleh ahli yang memiliki kualifikasi dan pengalaman relevan. Ahli pelabelan menggunakan acuan yang jelas untuk mengkategorikan teks *tweet* ke dalam kelas-kelas yang telah ditentukan. Visualisasi distribusi sentimen pada Gambar 3.1 menunjukkan dataset terdiri dari 2113 *tweet* total, dengan 1485 *tweet* (70.3%) berlabel negatif dan 628 *tweet* (29.7%) berlabel positif. Distribusi ini mengindikasikan ketidakseimbangan kelas yang signifikan, dimana sentimen negatif jauh lebih dominan. Meskipun terdapat ketidakseimbangan besar, jumlah data ini masih cukup representatif untuk klasifikasi teks terkait topik Quick Count Pemilihan Presiden Indonesia 2024 pada platform X Twitter. Hasil proses *labeling* data dapat dilihat pada Gambar 3.1, yang menampilkan distribusi sentimen dalam bentuk pie chart.



**Gambar 3.1** Tampilan Hasil Distribusi Sentimen

### 3.3 Pre-Processing

*Preprocessing* adalah tahap yang krusial dalam penelitian untuk memastikan data teks siap digunakan dalam proses klasifikasi teks. Berikut adalah masing-masing tahapan *preprocessing* yang telah dilakukan:

#### 3.3.1 Case folding

Pada tahap ini, seluruh teks dalam kolom *full\_text* diubah menjadi huruf kecil menggunakan fungsi *case\_folding*. Hasilnya dapat dilihat pada kolom '*case\_folding*' di output. Proses ini berhasil mengkonversi semua karakter menjadi huruf kecil, memastikan konsistensi format teks untuk tahapan selanjutnya. Berikut adalah tabel hasil dari tahap *case\_folding* pada beberapa contoh teks:

**Tabel 3.2** Hasil *Case folding*

No	full_text	case_folding
1	@m4retha Bilas muka gosok gigi liat quick count lagi	@m4retha bilas muka gosok gigi liat quick count lagi
2	@nvtamarsaoly @convomfs Bener bgt. Ga jauh2 hasil real count nanti sama quick count. Udah nder capee ngomong sama yang ga mau kalah sipaling dicurangi dan tersakiti .	@nvtamarsaoly @convomfs bener bgt. ga jauh2 hasil real count nanti sama quick count. udah nder capee ngomong sama yang ga mau kalah sipaling dicurangi dan tersakiti .
3	Yg udh liat dirty vote masa percaya ma Quick Count .. Ayokk semangat para pejuang amin Real count masih berjalan	yg udh liat dirty vote masa percaya ma quick count .. ayokk semangat para pejuang amin real count masih berjalan
...	...	...
2111	@andre_rosiade @prabowo @gibran_tweet Quick count hasil resmi kah? Lgsg lantik aja besok abis subuh	@andre_rosiade @prabowo @gibran_tweet quick count hasil resmi kah? lgsg lantik aja besok abis subuh
2112	@tanyarlfs Perasaan itu bukan pidato kemenangan deh tadi gw nonton fyi aja itu tuh emang udah direncanain kali acaranya buat ngelihat hasil quick count bareng timsesnya plis deh jangan kemakan judul berita minimal tonton sampe habis	@tanyarlfs perasaan itu bukan pidato kemenangan deh tadi gw nonton fyi aja itu tuh emang udah direncanain kali acaranya buat ngelihat hasil quick count bareng timsesnya plis deh jangan kemakan judul berita minimal tonton sampe habis
2113	@m4retha Update 23.29: 02 unggul dengan suara diatas 50% pada 29 provinsi. Total quickcount di 38 provinsi + Luar Negeri: 01 Amin 32 50% 02 Pagi 51 21% 03 Gama 16 29%	@m4retha update 23.29: 02 unggul dengan suara diatas 50% pada 29 provinsi. total quickcount di 38 provinsi + luar negeri: 01 amin 32 50% 02 pagi 51 21% 03 gama 16 29%

Proses case folding berhasil mengubah seluruh teks menjadi huruf kecil, yang penting untuk menyeragamkan format teks. Hal ini membantu mengurangi dimensi data dan meningkatkan konsistensi.

### 3.3.2 *Cleansing*

Tahap pembersihan teks dilakukan menggunakan fungsi *clean\_text* yang menghapus berbagai elemen tidak penting. Hasil *cleansing* terlihat pada kolom '*cleansing*', di mana teks menjadi lebih bersih dan hanya berisi kata-kata yang relevan untuk proses analisis. Berikut adalah tabel hasil dari tahap *cleansing* pada beberapa contoh teks:

**Tabel 3.3** Hasil *Cleansing*

No	case_folding	cleansing
1	@m4retha bilas muka gosok gigi liat quick count lagi	bilas muka gosok gigi liat quick count lagi
2	@nvtamarsaoly @convomfs bener bgt. ga jauh2 hasil real count nanti sama quick count. udah nder capee ngomong sama yang ga mau kalah sipaling dicurangi dan tersakiti .	bener bgt ga jauh hasil real count nanti sama quick count udah nder capee ngomong sama yang ga mau kalah sipaling dicurangi dan tersakiti
3	yg udh liat dirty vote masa percaya ma quick count .. ayokk semangat para pejuang amin real count masih berjalan	yg udh liat dirty vote masa percaya ma quick count ayokk semangat para pejuang amin real count masih berjalan
...	...	...
2111	@andre_rosiade @prabowo @gibran_tweet quick count hasil resmi kah? lgsg lantik aja besok abis subuh	quick count hasil resmi kah lgsg lantik aja besok abis subuh
2112	@tanyarlfs perasaan itu bukan pidato kemenangan deh tadi gw nonton fyi aja itu tuh emang udah direncanain kali acaranya buat ngelihat hasil quick count bareng timsesnya plis deh jangan kemakan judul berita minimal tonton sampe habis	perasaan itu bukan pidato kemenangan deh tadi gw nonton fyi aja itu tuh emang udah direncanain kali acaranya buat ngelihat hasil quick count bareng timsesnya plis deh jangan kemakan judul berita minimal tonton sampe habis
2113	@m4retha update 23.29: 02 unggul dengan suara diatas 50% pada 29 provinsi. total quickcount di 38 provinsi + luar negeri: 01 amin 32 50% 02 pagi 51 21% 03 gama 16 29%	update unggul dengan suara diatas pada provinsi total quickcount di provinsi luar negeri amin pagi gama

Proses *cleansing* berhasil menghapus elemen-elemen yang tidak penting seperti username Twitter, tanda baca, dan karakter khusus. Ini membantu mengurangi noise dalam data dan memfokuskan analisis pada konten teks yang substansial.

### 3.3.3 Stopword Removal

Penghapusan *stopwords* dilakukan menggunakan fungsi *remove\_stopwords*, yang menghilangkan kata-kata umum yang tidak signifikan dalam analisis. Daftar *stopwords* yang digunakan mencakup kata-kata baku bahasa Indonesia serta tambahan kata-kata informal. Hasil proses ini terlihat pada kolom '*stopword\_removal*', di mana kata-kata yang tersisa adalah kata-kata yang lebih bermakna untuk analisis. Berikut adalah tabel hasil dari tahap *stopword removal* pada beberapa contoh teks:

**Tabel 3.4** Hasil *Stopword Removal*

No	cleansing	stopword_removal
1	bilas muka gosok gigi liat quick count lagi	bilas muka gosok gigi liat quick count
2	bener bgt ga jauh hasil real count nanti sama quick count udah nder capee ngomong sama yang ga mau kalah sipaling dicurangi dan tersakiti	bener hasil real count quick count capee ngomong kalah sipaling dicurangi tersakiti
3	yg udh liat dirty vote masa percaya ma quick count ayokk semangat para pejuang amin real count masih berjalan	liat dirty vote percaya quick count ayokk semangat pejuang amin real count berjalan
...	...	...
2111	quick count hasil resmi kah lgsg lantik aja besok abis subuh	quick count hasil resmi lantik besok subuh
2112	perasaan itu bukan pidato kemenangan deh tadi gw nonton fyi aja itu tuh emang udah direncanain kali acaranya buat ngelihat hasil quick count bareng timsesnya plis deh jangan kemakan judul berita minimal tonton sampe habis	pidato kemenangan nonton emang direncanain acaranya hasil quick count bareng timsesnya kemakan judul berita minimal tonton sampe habis
2113	update unggul dengan suara diatas pada provinsi total quickcount di provinsi luar negeri amin pagi gama	update unggul suara diatas provinsi total quickcount provinsi negeri amin

Proses *stopword removal* berhasil menghilangkan kata-kata umum yang tidak signifikan dalam analisis, membantu mengurangi dimensi data dan memfokuskan analisis pada kata-kata yang lebih bermakna. Penggunaan daftar *stopwords* yang mencakup kata-kata baku dan informal bahasa Indonesia merupakan langkah yang baik. Namun, terdapat inkonsistensi dalam penghapusan *stopwords*. Beberapa kata informal atau slang tidak terdeteksi sebagai *stopwords*, contohnya seperti "capee" dan "sipaling" pada baris ke-2, serta "emang" pada baris ke-2112. Kata-kata ini merupakan bentuk informal yang umumnya dianggap sebagai *stopword* dan idealnya harus dihilangkan dalam proses *stopword removal*.

### 3.3.4 Stemming

Tahap terakhir melibatkan proses stemming menggunakan fungsi *stem\_text* dengan bantuan *library Sastrawi*. Proses ini mengubah kata-kata menjadi bentuk dasarnya. Hasil stemming ditampilkan pada kolom '*stemming*', di mana kata-kata telah disederhanakan ke bentuk akarnya. Ini membantu mengurangi variasi kata dan meningkatkan efektivitas analisis. Berikut adalah tabel hasil dari tahap *stemming* pada beberapa contoh teks:

**Tabel 3.5 Hasil Stemming**

No	stopword_removal	stemming
1	bilas muka gosok gigi liat quick count bener hasil real count quick count	bilas muka gosok gigi liat quick count bener hasil real count quick count
2	capee ngomong kalah sipaling dicurangi tersakiti liat dirty vote percaya quick count	capee ngomong kalah sipaling curang sakiti liat dirty vote percaya quick count
3	ayokk semangat pejuang amin real count berjalan	ayokk semangat juang amin real count jalan
...	...	...
2111	quick count hasil resmi lantik besok subuh	quick count hasil resmi lantik besok subuh
2112	pidato kemenangan nonton emang direncanain acaranya hasil quick count bareng timsesnya kemakan judul berita minimal tonton sampe habis	pidato menang nonton emang direncanain acara hasil quick count bareng timsesnya makan judul berita minimal tonton sampe habis
2113	update unggul suara diatas provinsi total quickcount provinsi negeri amin	update unggul suara atas provinsi total quickcount provinsi negeri amin

Proses stemming dengan menggunakan *library Sastrawi* berhasil mengubah kata-kata menjadi bentuk dasarnya, membantu mengurangi variasi kata dan meningkatkan efektivitas analisis. Meskipun demikian, terdapat beberapa kekurangan dalam hasil stemming. Beberapa kata tidak di-stemming dengan benar, seperti "direncanain" pada baris ke-2112 yang seharusnya menjadi "rencana".

### 3.4 Visualisasi

Visualisasi *Word Cloud* yang ditampilkan pada Gambar 3.2 memberikan gambaran yang jelas tentang kata-kata dalam setiap kategori sentimen. Untuk sentimen negatif, kata-kata yang menonjol meliputi "curang", "real", "hasil", dan "kpu", mengindikasikan fokus pada isu-isu seputar kecurangan dan hasil penghitungan suara. Sementara itu, *Word Cloud* sentimen positif

menampilkan kata-kata seperti "kpu", "suara", "hasil", dan "menang" sebagai yang paling dominan, menunjukkan kecenderungan membahas kepercayaan terhadap proses pemilu dan hasil yang menguntungkan.



Gambar 3.2 Tampilan Hasil *Word Cloud*

### 3.5 Pembobotan Kata

Berdasarkan hasil pembobotan TF-IDF yang ditampilkan pada gambar 3.3, setiap *term* dalam dataset telah diubah menjadi vektor numerik. Total kosakata yang dianalisis adalah 4999 *term*. Pada *output* (0, 838) yang menunjukkan bahwa pada dokumen ke-0 terdapat *term* pada *index* ke-838 dengan nilai 0,0688. Untuk *output* (2112, 146) menunjukkan bahwa pada dokumen ke-2112 terdapat *term* pada *index* ke-146 dengan nilai 0,2451. Proses ini menghasilkan matriks TF-IDF yang mencakup 2112 dokumen dan 4999 *term*.

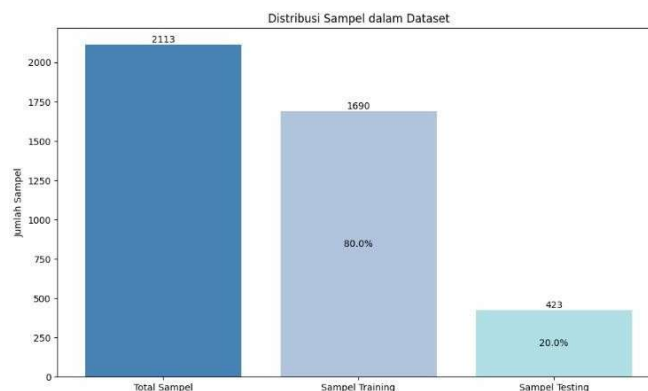
Total vocabulary: 4999	
Term	Frequency-Inverse Document Frequency (TF-IDF):
(0, 838)	0.06886996842891753
(0, 3669)	0.06886996842891753
(0, 2466)	0.23852635668016155
(0, 1458)	0.45435830847651
(0, 1494)	0.45435830847651
(0, 2868)	0.4516208951228151
(0, 536)	0.5109408293786471
(1, 3861)	0.3708709811925937
(1, 888)	0.22228298442708463
(1, 4189)	0.46581517195266436
(4, 2801)	0.235945860952625
(1, 3067)	0.39168913667026514
(1, 716)	0.4908067348679603
(1, 3720)	0.1733845454710436
(4, 1583)	0.11846268284619828
(1, 449)	0.28542768295740477
(1, 838)	0.11244662346017476
(4, 3669)	0.068231173908778
(2, 1841)	0.3231369472829371
(2, 146)	0.2701786628891291
(2, 1950)	0.2363684743678393
(2, 4820)	0.2551545669615561
(2, 295)	0.46615559856142
(2, 3411)	0.19556787121704372
(2, 4820)	0.3876288689895403
(2, 1054)	0.37784708921422624
(2, 3720)	0.17184915083145502
(2, 836)	0.111386721180153
(2, 3669)	0.06565436856515705
(2, 2466)	0.2273893219790304
(3, 1129)	0.329385079990821
(3, 987)	0.3821861212629686
(3, 2414)	0.36264456799218665
(3, 4851)	0.369992952672523816
(3, 3461)	0.2490621692789843
(3, 350)	0.2339495746235517
(3, 1432)	0.3821001212629686
(3, 716)	0.2578564866668387
(3, 3461)	0.28321657274917
(3, 3411)	0.3871986166874691
...	...
(2112, 3670)	0.199555639791592
(2112, 4317)	0.15476864059084822
(2112, 146)	0.2451273162975915

Gambar 3.3 Tampilan Hasil Pembobotan Kata

Matriks TF-IDF yang dihasilkan menggambarkan bobot setiap term dalam setiap dokumen. Nilai-nilai dalam matriks ini mencerminkan tingkat kepentingan suatu term dalam dokumen tertentu relatif terhadap seluruh kumpulan dokumen. Visualisasi matriks ini memungkinkan untuk melihat penyebaran bobot term di seluruh kumpulan dokumen, membantu dalam identifikasi pola dan tema utama dalam kumpulan dokumen.

### 3.6 *Split Data*

Proses *split* data menghasilkan pembagian dataset menjadi dua bagian utama, dengan 1690 sampel (80%) digunakan sebagai data *training* dan 423 sampel (20%) sebagai data *testing* dari total 2113 sampel. Pembagian dengan rasio 80:20 ini dipilih untuk memberikan keseimbangan optimal antara jumlah data yang cukup untuk melatih model *Naive Bayes* dan data yang memadai untuk menguji performa model secara akurat. Visualisasi *bar chart* memperlihatkan distribusi sampel dengan jelas, menampilkan perbandingan antara total sampel, sampel *training*, dan sampel *testing* beserta persentasenya. Hasil pembagian data ini secara visual direpresentasikan dalam Gambar 3.4, yang menampilkan distribusi sampel dalam dataset.



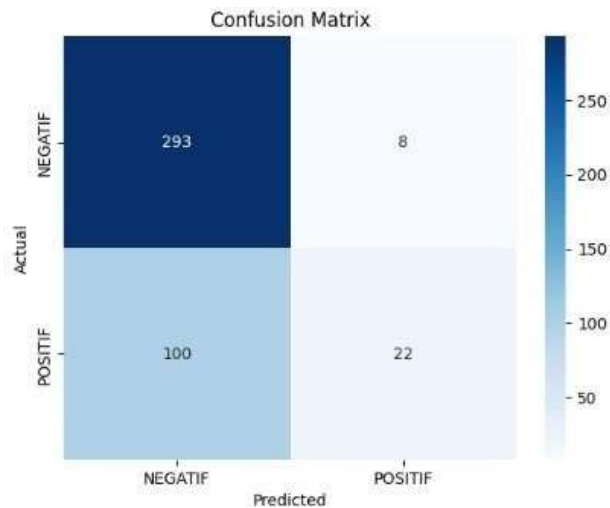
**Gambar 3.4** Tampilan Hasil *Split Data*

### 3.7 Klasifikasi

Hasil analisis menggunakan model klasifikasi *Naive Bayes* juga ditampilkan dalam bentuk confusion matrix pada Gambar 3.5. Matriks ini memberikan gambaran detail tentang performa model dalam memprediksi sentimen. Matriks menunjukkan bahwa model berhasil



memprediksi 293 kasus sebagai *true negative* dan 22 kasus sebagai *true positive*. Namun, terdapat 100 kasus diklasifikasikan sebagai *false negative* dan 8 kasus diklasifikasikan sebagai *false positive*. Visualisasi ini membantu dalam memahami distribusi prediksi model secara lebih mendalam.



**Gambar 3.5** Tampilan Hasil *Confusion matrix* Metode *Naive Bayes*

### 3.8 Evaluasi

Untuk memvalidasi hasil akurasi yang diperoleh dari model, dapat melakukan perhitungan manual menggunakan rumus akurasi berdasarkan data dari *confusion matrix*. Akurasi model dapat dihitung (3.1) sebagai berikut:

$$Accuracy = \frac{22 + 293}{22 + 100 + 8 + 293} = \frac{315}{423} = 0,7446 \quad (3.1)$$

Hasil akurasi berdasarkan data dari *confusion matrix*, yaitu 74,46%. Perhitungan manual ini memverifikasi keakuratan hasil yang diberikan oleh model klasifikasi *Naive Bayes*. Meskipun hasil ini menunjukkan performa yang cukup baik, masih ada ruang untuk peningkatan, misalnya dengan mengoptimalkan fitur, menambah data atau mencoba algoritma klasifikasi lainnya.