

BAB IV HASIL DAN PEMBAHASAN

4.1 Hasil Scrapper Data

Pengambilan data dari internet atau yang lebih dikenal dengan *scraping* telah menjadi alat yang sangat berguna dalam mengumpulkan informasi secara otomatis dari berbagai sumber *online*. Dalam konteks politik, informasi tentang calon presiden dan wakil presiden sangatlah penting, baik untuk analisis politik maupun pemantauan publik.

Scrapper data didapat dari TikTok yang diambil dari halaman *fanpage* aplikasi TikTok tentang capres dan cawapres tahun 2024. Data komentar diperoleh dari *fanpage* TikTok yang terlihat pada tabel di bawah ini.

Tabel 4.1 Data *Crawling/Scrapper*

No	Halaman Fanpage/URL
1	https://www.tiktok.com/@zeinchannel410/video/7267127149600017670
2	https://www.tiktok.com/@metro_tv/video/7306089644578360578
3	https://www.tiktok.com/@narasi/video/7281587109260201221
4	https://www.tiktok.com/@heerrzyy/video/7154497210330254618
5	https://www.tiktok.com/@undergreen.id/video/7269349287924993285
6	https://www.tiktok.com/@hanapijani/video/7293100060034764037
7	https://www.tiktok.com/@hanapijani/video/7291177260223089926

Halaman *link* data tersebut nantinya di *copy* pada folder “data” yang telah di siapkan. *Script* yang akan dibuat akan memanfaatkan pustaka-pustaka *Python* seperti untuk mengambil dan memproses data dari halaman TikTok yang berisi informasi tentang komentar capres dan cawapres 2024. Hasil dari *scraping* akan disimpan dalam format XLXS agar mudah diakses dan dianalisis lebih lanjut. Adapun *script* pengambilan data komentar dari halaman TikTok terlihat pada

gambar di bawah ini. Tahapan pengambilan data komentar TikTok adalah sebagai berikut :

- a. Lakukan pemilihan *link* atau *fanpage* TikTok yang akan diambil data komentarnya.
- b. Copy URL ke dalam folder yang telah disediakan.
- c. Jalankan scrapper dan processing.bat dengan cara double klik.
- d. Sistem akan menjalankan pengambilan data komentar TikTok dan nantinya akan data komentar tersebut dimasukkan dalam media excel dengan format xlxs.

```

38 # Mengambil data/komentar dari tiktok (Scraper)
39 async with TikTokApi() as api:
40     await api.create_sessions(ms_tokens=[ms_token], num_sessions=1, sleep_after=10, headless=False)
41     video = api.video(id=videoid)
42
43     comments = []
44
45     async for comment in video.comments(count=250):
46         comments.append([
47             datetime.fromtimestamp(comment.as_dict['create_time']).strftime("%Y-%m-%d"),
48             datetime.fromtimestamp(comment.as_dict['create_time']).strftime("%H:%M:%S"),
49             comment.as_dict['text'], comment.as_dict['user']['nickname']
50         ])
51
52         with open("data/raw_" + videoid + ".txt", 'w') as convert_file:
53             convert_file.write(json.dumps(comment.as_dict))
54
55     df = pd.DataFrame(comments)
56
57     # Simpan dataframe ke excel
58     namefile = videoid + ".xlsx"
59     df.to_excel("data/" + namefile, index=False, header=["Tanggal", "Waktu", "Text", "User"])
60
61     # Tambahkan daftar nama file
62     with open("namefiles", "a") as fl:
63         fl.write(namefile)
64         fl.write("\n")
65

```

Gambar 4.1 Script Scrapper/Crawling

Tanggal	Waktu	Text	User
2024-02-05	10:03:16	02 ya pasti tetap Prabowo lah	Budak kjangmV
2023-10-25	15:48:41	prabowo presiden RI 2024 pilihan saya	paulinanompetustD
2023-11-27	23:30:02	nomor 1	maudythalita
2024-01-25	11:14:20	Ganjar mahfud	syaddamsentaru
2023-11-14	21:01:29	selamat nomor urut anda benar	Mohammad Ridho
2023-11-02	19:30:45	@Mdaa se teppak nomor brmpa rahh	v
2023-12-21	10:31:38	yang genius pilih Prabowo Gibran	anza_lina27
2023-12-15	08:35:03	Prabowo dan Gibran	username
2024-01-10	16:35:34	Prabowo Gibran	Elyaningtyas
2023-12-30	11:44:51	lu aminin semua bg	Rell bae
2023-10-26	09:30:01	Anies-imin sobat hadir	Kontrol
2023-12-17	11:39:20	Bismillahirrahmanirrahim saya Prabowo	Neng Rina
2023-12-06	16:23:21	ganjar	sghrp?
2023-12-26	21:09:06	GAMA	Umar Mukhtar
2023-12-20	06:49:00	indonesia butuh presiden yang ber pengalaman dalam segala hal dan bisa mengerti keadaan rakyat nya yang susah semangat jangan sampai salah pilih yaa	Eu sp??
2023-12-14	10:46:49	gibran	Nerti Yunita Sari
2023-11-05	21:20:12	anislah yg terbaik	md azwentam1974
2023-12-01	19:47:48	psi ganjar?	malpin still xp
2023-12-13	08:14:34	prabowo gibran	Suka Wiyang
2023-11-04	21:00:21	pak Anis cakimin yess good	@NeNe' lTeNg
2024-02-05	10:03:35	pokoknya Prabowo	Budak kjangmV
2023-12-12	22:03:33	PRABOWO	mahfudzariant095
2023-11-05	20:30:04	Pak Anies lah	Ciki
2023-11-15	13:47:35	jelas no 3 lah	No nickname
2024-02-02	01:54:00	anis baswedan	syafjintingshuka
2024-02-05	20:52:34	Perasaan, paslon nomor 02 tidak begitu potonya di kertas suara..	Ujang Emka
2023-11-20	18:09:10	bukannya PSI all in Prabowo ya?	ty
2023-10-31	23:03:38	pemilik akunnnya lagi milih prabowo	putra_Qalbi
2023-12-23	10:22:36	Prabowo-gibran	Mei...
2024-02-02	08:08:30	tetap prabowo gibran is the best	IrmakalLa
2023-11-05	14:15:51	tetap BP anis	Eesolih
2024-01-26	11:16:20	ya jelas dua lah	alkijamet
2024-01-21	13:30:22	harus nya silih PSI dukung pak prabowo-gibran soalnya ketua umum PSI itu Kaesang adik nya pak gibran	فاهري باگوس فبرمانشه
2024-01-08	20:33:52	leh bukannya psi itu dukung prabowo gibran yaa?	fazam
2024-01-16	07:16:29	bapak Kita donk PAK PRABOWO	Hanan beauty HNB
2024-01-04	16:14:04	komen ya banyak pendukung prabowogibransaya jawa timur juga prabowo gibran	Ahmad Naufal
2024-01-11	17:35:49	lah psi ke ganjar?	jan

Gambar 4.2 Data Komentar Hasil *Scraper/Crawling* Capres dan Cawapres TikTok

4.2 Hasil Preprocessing Data

Preprocessing dilakukan pada tiap baris data pada kolom komentar. *Preprocessing* terdiri dari *case folding*, *removing punctuation*, *removing unnecessarily whitespace*, *tokenizing*, dan *filtering*. Data komentar TikTok pada analisis sentimen capres dan cawapres 2024 yang telah didapat dari tahap *scraper*, kemudian di lakukan *preprocessing* tiap baris pada data komentar dengan tahapan berikut :

a. Case Folding

Merupakan tahapan di mana semua huruf pada teks diubah menjadi huruf kecil. Ini membantu dalam konsistensi teks, sehingga "Kucing" dan "kucing" dianggap sama dalam analisis teks. Adapun hasil data komentar yang telah melalui tahapan *case folding* dari *preprocessing* terlihat pada gambar di bawah ini.

Tanggal	Waktu	Text	User	Casefolding
2023-08-11	18:04:22	bpk prabowo subianto presiden ku 2024 ri	Logan	bpk prabowo subianto presiden ku 2024 ri
2023-08-11	18:20:44	prabowo presiden gaspoli prabowo kasih kendur	budyanto387	pak prabowo presiden. gaspoli pak prabowo jangan kasih kendur
2023-08-11	16:03:11	prabowo	darfitahtiar	prabowo
2023-08-11	22:02:45	apapun kualitasmu tetep ganjar presiden	qiaramalastari	sebesar apapun kualitasmu tetep ganjar presiden
2023-08-11	14:21:28	prabowo ri 1	ulive_20	prabowo ri 1
2023-08-11	16:02:59	prabowo subianto pasti menang di pilpres 2024	user5875346756178ibnu manan	pak prabowo subianto pasti menang di pilpres 2024
2023-08-11	15:48:15	praboworidwan kamil	@JayAryam07	prabowo-ridwan kamil
2023-08-11	20:20:12	jitra alah ktp mu pilih prabowo	mitra	@jitra... alah ktp mu, kita pilih prabowo
2023-08-11	15:41:42	prabowo 2024	syahfitriandhan4	prabowo-2024
2023-09-01	13:49:44	tangerang barten pilih pa anis	ibu ade	siapa saya, dari, tangerang, barten, siap, pilih, pa, anis
2023-08-11	00:11:16	ganjar	Dada rosidah	pak ganjar
2023-08-11	19:52:18	anis mantap	viky	anis mantap
2023-08-11	17:52:44	prabowo	lucky	prabowo
2023-08-11	20:13:51	masyarakat butuh perubahan	tang subroto	masyarakat butuh perubahan
2023-08-11	19:22:45	pa ganjar	jihanis	pa ganjar
2023-08-11	03:58:05	anies pilih rakyat no wowo no uban	suradaja911	anies di pilih rakyat no wowo, no uban
2023-08-11	01:36:59	ganjar pilihanku	jonindra15	pak ganjar pilihanku
2023-08-11	14:33:40	prabowo dukung sangkum buhar	SWAETR 226	prabowo di dukung sangkum buhar buhar
2023-08-11	22:58:39	pilih prabowo males golkar pan gara2 si artis pelakor shahnaz nyalon	bau momok	mau pilih prabowo males ada golkar sama pan gara2 si artis pelakor shahnaz nyalon
2023-09-01	18:06:24	klatan jawa dukung anis baswedan cak imin capres cawapres 2024	Bejo Alexander	aku klatan jawa tengah dukung anis baswedan dan cak imin menuju capres cawapres 2024
2023-09-2	14:51:36	ign banteng diatas	daniel	ign ada lagi banteng diatas
2023-08-11	22:45:35	jateng tetep milih bp prabowo	wim	jateng tetep milih bp prabowo
2023-08-31	11:70:00	prabowo	Arjuno24	prabowo
2023-08-31	16:41:51	not hard feeling ya udih dipimpin partai pdi kalo sih pgn yg beda ya jd pgn gantet	Provisionadine	not hard feeling ya, cuma kalian kan udih pernah dipimpin dari partai pdi, kalo aku sih pgn sesuatu yg beda ya jd pgn gantet
2023-08-11	11:13:40	bekasi full prabowo	Bob_08	bekasi full prabowo
2023-08-31	16:56:00	hanura yg punya tv balikan fakta berita	Blksky7	hanura masalah yg punya tv bisa di balikan fakta berita
2023-08-11	18:51:06	prabowo	yulder_@M99	prabowo
2023-08-11	10:32:53	prabowo	rudhy093	prabowo
2023-08-11	17:13:05	insya allah anis	parwito45	insya allah pak anis
2023-08-11	17:17:34	anies	AnakStampal	anies
2023-08-11	14:41:03	aniessssssss yessss	asari093	pak aniessssss yessss
2023-08-11	19:27:20	pk anis	Fathorrahman_04	pk anis
2023-08-11	15:05:34	anies yesss	Ahmad R	anies yesss
2023-08-11	11:58:36	aniuh yg menang	cucu sumbar	aniuh yg menang
2023-08-11	18:03:15	anies rasid baswedan presiden	h_nur hasanah	pak anies rasid baswedan presiden
2023-08-11	17:40:03	anies	Alyasir593	anies
2023-08-11	14:30:48	pan pan pan ganjar	Lot Tik	pan pan pan ganjar semakin di depan

Gambar 4. 3 Tahap *Case Folding*

Penjelasan detail mengenai *case folding* adalah sebagai berikut :

1. Merubah Huruf kecil menjadi Huruf Besar atau sebaliknya

Dalam tahapan ini, setiap huruf dalam teks diubah menjadi huruf kecil atau huruf besar, sesuai dengan preferensi atau kebutuhan analisis. Misalnya, teks "Hello World" akan diubah menjadi "hello world" jika menggunakan *case folding* untuk mengubah ke huruf kecil.

2. Mengatasi Inkonsistensi dalam Penulisan Huruf

Dalam banyak kasus, data teks yang dikumpulkan dari berbagai sumber dapat memiliki inkonsistensi dalam penulisan huruf, seperti beberapa kata ditulis dengan huruf kecil dan beberapa kata lainnya ditulis dengan huruf besar. Dengan menggunakan *case folding*, inkonsistensi semacam itu dapat diatasi, sehingga memastikan bahwa semua kata dalam teks ditulis dengan format yang seragam.

3. Meningkatkan Efisiensi Pemrosesan Teks

Dengan mengubah semua huruf dalam teks menjadi huruf kecil atau huruf besar, kita dapat menghindari masalah kesalahan pencocokan kata yang disebabkan oleh perbedaan penulisan huruf. Misalnya, pencocokan kata "Hello" dengan "hello" akan berhasil setelah dilakukan *case folding*, sehingga meningkatkan efisiensi dalam pemrosesan teks.

4. Mengurangi Variasi Kata

Dalam beberapa kasus, *case folding* juga dapat membantu dalam mengurangi variasi kata, terutama jika kita tertarik pada analisis yang lebih fokus pada konten kata daripada format penulisannya. Dengan mengubah semua kata menjadi huruf kecil, kita dapat memperlakukan kata-kata yang sama yang ditulis dengan huruf besar dan huruf kecil sebagai kata yang sama dalam analisis.

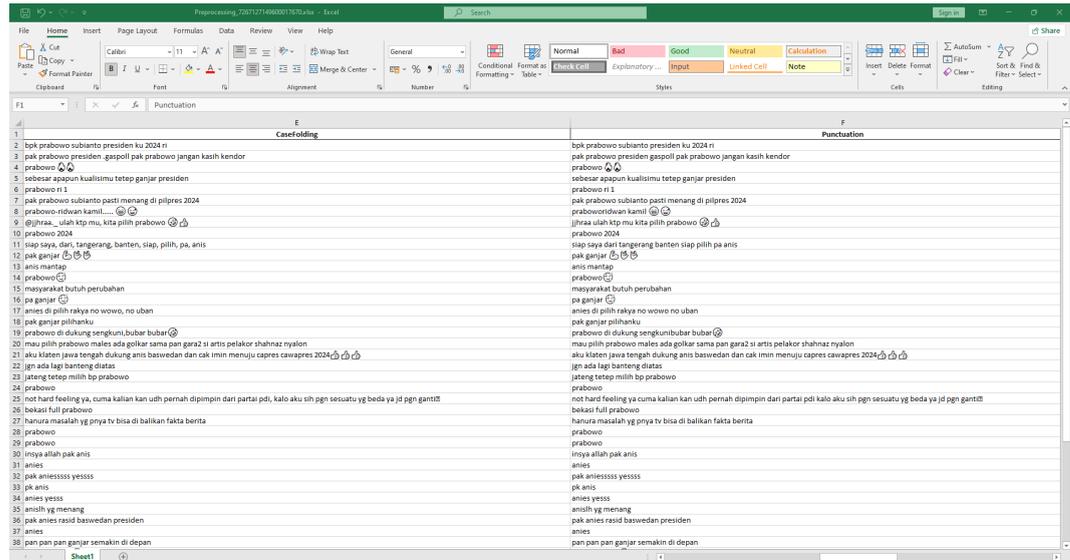
Dengan menggunakan *case folding* dalam tahapan *preprocessing* teks, kita dapat memastikan bahwa data teks yang akan diolah menjadi lebih konsisten dan siap untuk analisis lebih lanjut, termasuk analisis sentimen, klasifikasi teks, atau pemodelan bahasa. Adapun *script* untuk menampilkan *case folding* setelah di dapat data komentar hasil *scraper/crawling* adalah `v = v.lower()`. Setiap huruf dalam teks ‘v’ diubah menjadi huruf kecil semua menggunakan metode `lower()`. Ini bertujuan untuk memastikan konsistensi dalam penulisan huruf, sehingga mempermudah analisis teks tanpa memperhatikan perbedaan besar-kecil huruf

b. *Removing Punctuation*

Pada tahap ini, tanda baca seperti titik, koma, tanda seru, dan lain sebagainya dihapus dari teks. Ini membantu fokus pada kata-kata kunci dan menghilangkan gangguan yang tidak relevan. *Script* yang digunakan adalah :

`‘v’ = v.translate(str.maketrans("", "", string.punctuation))`.

Menggunakan metode `translate()` untuk menghapus semua tanda baca dari teks v. Ini dilakukan dengan membuat tabel translasi kosong yang memerintahkan penghapusan semua karakter tanda baca. Adapun hasil data komentar yang telah melalui tahapan *removing punctuation* dari *preprocessing* terlihat pada gambar di bawah ini.



Gambar 4. 4 Tahap *Removing Punctuation*

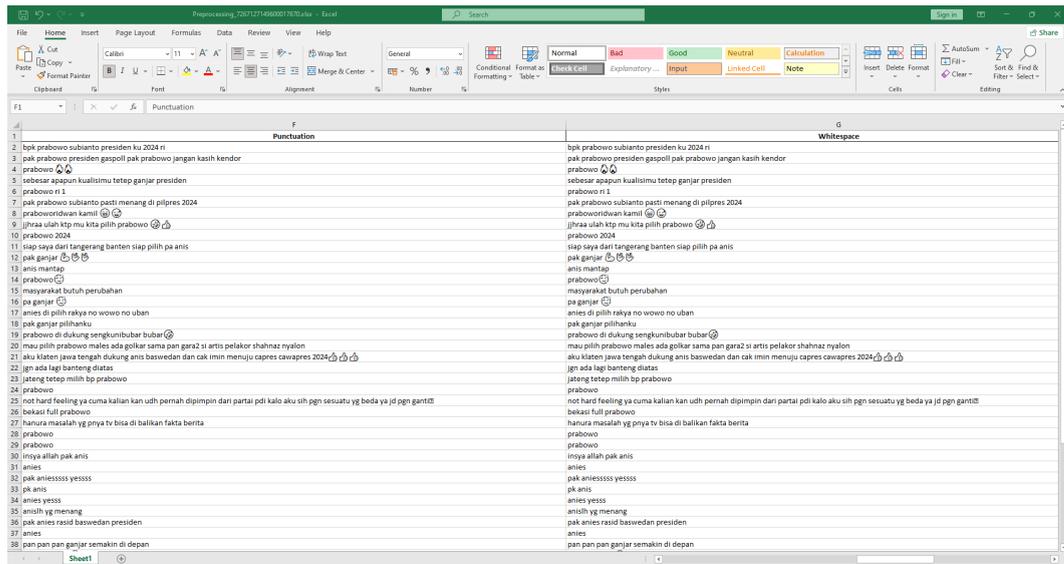
c. *Removing Unnecessarily Whitespace*

Whitespace yang tidak diperlukan, seperti spasi ganda atau tab yang berlebihan, dihapus. Ini membantu memperbaiki format teks dan membuatnya lebih mudah untuk diproses. *Script* yang digunakan adalah :

$$v = v.\text{strip}()$$

$$v = \text{re.sub}(r"\s+", " ", v)$$

Tahap ini dilakukan dalam dua langkah. Pertama, menggunakan metode **strip()** untuk menghapus *whitespace* yang tidak perlu di awal dan akhir teks. Kemudian, menggunakan ekspresi reguler **re.sub()** untuk mengganti semua urutan *whitespace* yang berulang (termasuk spasi ganda) dengan satu spasi tunggal. Ini membantu memastikan format teks yang konsisten dan terstruktur. Adapun hasil data komentar yang telah melalui tahapan *removing unnecessarily whitespace* dari *preprocessing* terlihat pada gambar di bawah ini.



Gambar 4. 5 Tahap *Removing Unnecessarily Whitespace*

d. *Tokenizing*

Tokenizing adalah proses membagi teks menjadi unit-unit kecil yang disebut token. Token bisa berupa kata, frasa, atau simbol tertentu. Biasanya, dalam konteks NLP (Natural Language *Processing*), tokenisasi dilakukan dengan membagi teks menjadi kata-kata. *Script* yang digunakan adalah :

```
tokenwords = nltk.tokenize.word_tokenize(v)
```

Tahap ini menggunakan modul **nltk** untuk membagi teks menjadi token-token (kata-kata) menggunakan metode **word_tokenize()**. Setiap token akan merepresentasikan sebuah kata dalam teks Adapun hasil data komentar yang telah melalui tahapan *tokenizing* dari *preprocessing* terlihat pada gambar di bawah ini.

Whitespace	Tokenizing
1	
2 bpk prabowo subianto presiden ku 2024 r1	['bpk', 'prabowo', 'subianto', 'presiden', 'ku', '2024', 'r1']
3 pak prabowo presiden gaspol pak prabowo jangan kasih kendor	['pak', 'prabowo', 'presiden', 'gaspol', 'pak', 'prabowo', 'jangan', 'kasih', 'kendor']
4 prabowo	['prabowo']
5 sebesar apapun kualitasmu tetep ganjar presiden	['sebesar', 'apapun', 'kualitasmu', 'tetep', 'ganjar', 'presiden']
6 prabowo r1	['prabowo', 'r1']
7 pak prabowo subianto pasti menang di pilpres 2024	['pak', 'prabowo', 'subianto', 'pasti', 'menang', 'di', 'pilpres', '2024']
8 prabowordihan kami	['prabowordihan', 'kami']
9 jitraa adin ttp mu kita pilih prabowo	['jitraa', 'adin', 'ttp', 'mu', 'kita', 'pilih', 'prabowo']
10 prabowo 2024	['prabowo', '2024']
11 siap saya dari tangerang banten siap pilih pa anis	['siap', 'saya', 'dari', 'tangerang', 'banten', 'siap', 'pilih', 'pa', 'anis']
12 pak ganjar	['pak', 'ganjar']
13 anis mantap	['anis', 'mantap']
14 prabowo	['prabowo']
15 masyarakat butuh perubahan	['masyarakat', 'butuh', 'perubahan']
16 pa ganjar	['pa', 'ganjar']
17 anies di pilih rakyat no wowo no uban	['anies', 'di', 'pilih', 'rakyat', 'no', 'wowo', 'no', 'uban']
18 pak ganjar pilihanku	['pak', 'ganjar', 'pilihanku']
19 prabowo di dukung engkumihubar buhar	['prabowo', 'di', 'dukung', 'engkumihubar', 'buhar']
20 mau pilih prabowo males ada golkar sama pan gara2 si artis pelakor shahnaz nyalon	['mau', 'pilih', 'prabowo', 'males', 'ada', 'golkar', 'sama', 'pan', 'gara2', 'si', 'artis', 'pelakor', 'shahnaz', 'nyalon']
21 aku klaten jawa tengah dukung anis baswedan dan cak imin menuju capres cawapres 2024	['aku', 'klaten', 'jawa', 'tengah', 'dukung', 'anis', 'baswedan', 'dan', 'cak', 'imin', 'menuju', 'capres', 'cawapres', '2024']
22 gn ada lagi banteng diatas	['gn', 'ada', 'lagi', 'banteng', 'didas']
23 jateng tetep milih bp prabowo	['jateng', 'tetep', 'milih', 'bp', 'prabowo']
24 prabowo	['prabowo']
25 nih hard feeling ya cuma kalian kan udh pernah dipimpin dari partai pdi kalo aku sih pgn sesuatu yg beda ya jd pgn gentr	['nih', 'hard', 'feeling', 'ya', 'cuma', 'kalian', 'kan', 'udh', 'pernah', 'dipimpin', 'dari', 'partai', 'pdi', 'kalo', 'aku', 'sih', 'pgn', 'sesuatu', 'yg', 'beda', 'ya', 'jd', 'pgn', 'gentr']
26 bekas full prabowo	['bekas', 'full', 'prabowo']
27 hanura masalah yg prya tv bisa di balikan fakta berita	['hanura', 'masalah', 'yg', 'prya', 'tv', 'bisa', 'di', 'balikan', 'fakta', 'berita']
28 prabowo	['prabowo']
29 insya allah pak anis	['insya', 'allah', 'pak', 'anis']
31 anies	['anies']
32 pak anisesss yessss	['pak', 'aniesesss', 'yessss']
33 pk anis	['pk', 'anis']
34 anies yesss	['anies', 'yesss']
35 anisuh yg menang	['anisuh', 'yg', 'menang']
36 pak anies rasid baswedan presiden	['pak', 'anies', 'rasid', 'baswedan', 'presiden']
37 anies	['anies']
38 [pak pan pan ganjar semakin di depan	['pak', 'pan', 'pan', 'ganjar', 'semakin', 'di', 'depan']

Gambar 4. 6 Tahap Tokenizing

e. Filtering

Pada tahap ini, kita dapat memfilter kata-kata tertentu yang tidak relevan atau tidak diinginkan dari teks. Misalnya, kita bisa menghapus kata-kata yang terlalu umum (*stopwords*) atau kata-kata spesifik yang tidak relevan dengan analisis kita. *Script* yang digunakan adalah :

```
sw = set(stopwords.words("indonesian"))
```

```
filterwords = [t for t in tokenwords if t not in sw]
```

Tahap ini menghapus kata-kata tidak bermakna atau *stopwords* dari teks menggunakan kamus stopwords bahasa Indonesia dari modul **nltk**. Setiap token akan diverifikasi keberadaannya dalam kamus *stopwords*, dan hanya kata-kata yang tidak termasuk dalam *stopwords* yang akan dipertahankan. Adapun hasil data komentar yang telah melalui tahapan *filtering* dari *preprocessing* terlihat pada gambar di bawah ini.

Tokenizing	Filtering
[tok, 'prabowo, 'subianto, 'presiden, 'ku, '2024, 'y]	[tok, 'prabowo, 'subianto, 'presiden, 'ku, '2024, 'y]
[pa, 'prabowo, 'presiden, 'gaspol, 'pak, 'prabowo, 'jangan, 'kasih, 'kendor]	[prabowo, 'presiden, 'gaspol, 'prabowo, 'kasih, 'kendor]
[prabowo, 'y]	[prabowo, 'y]
[sebesar, 'apapun, 'kualitasmu, 'tetep, 'ganjar, 'presiden]	[apapun, 'kualitasmu, 'tetep, 'ganjar, 'presiden]
[prabowo, 'y, ']	[prabowo, 'y, ']
[pak, 'prabowo, 'subianto, 'pasti, 'menang, 'di, 'pilpres, '2024]	[prabowo, 'subianto, 'menang, 'pilpres, '2024]
[praboworduan, 'kami, ']	[praboworduan, 'kami, ']
[jihad, 'sah, 'sdp, 'mu, 'kita, 'pilih, 'prabowo, ']	[jihad, 'sah, 'sdp, 'mu, 'pilih, 'prabowo, ']
[prabowo, '2024]	[prabowo, '2024]
[siap, 'saya, 'dari, 'tangerang, 'banten, 'siap, 'pilih, 'pa, 'anis]	[tangerang, 'banten, 'pilih, 'pa, 'anis]
[pak, 'ganjar, ']	[ganjar, ']
[anis, 'menang]	[anis, 'menang]
[prabowo, ']	[prabowo, ']
[masyarakat, 'butuh, 'perubahan]	[masyarakat, 'butuh, 'perubahan]
[pak, 'ganjar, ']	[pak, 'ganjar, ']
[anis, 'di, 'pilih, 'takya, 'no, 'wowo, 'no, 'uban]	[anis, 'pilih, 'takya, 'no, 'wowo, 'no, 'uban]
[pak, 'ganjar, 'pilihanku]	[ganjar, 'pilihanku]
[prabowo, 'di, 'dukung, 'mengembalikan, 'bubar, ']	[prabowo, 'dukung, 'mengembalikan, 'bubar, ']
[mau, 'pilih, 'prabowo, 'males, 'ada, 'golkar, 'sama, 'pan, 'gara2, 'si, 'artis, 'pelakor, 'shahza, 'nyalon]	[pilih, 'prabowo, 'males, 'golkar, 'pan, 'gara2, 'si, 'artis, 'pelakor, 'shahza, 'nyalon]
[aku, 'klaten, 'jawa, 'tengah, 'dukung, 'anis, 'baswedan, 'dan, 'tak, 'lamin, 'menuju, 'capres, 'kawapres, '2024, ']	[klaten, 'jawa, 'dukung, 'anis, 'baswedan, 'tak, 'lamin, 'capres, 'kawapres, '2024, ']
[pni, 'ada, 'lagi, 'banteng, 'ditarai]	[pni, 'banteng, 'ditarai]
[jaten, 'tetep, 'milih, 'sp, 'prabowo]	[jaten, 'tetep, 'milih, 'sp, 'prabowo]
[prabowo]	[prabowo]
[not, 'hard, 'keeling, 'ya, 'cunai, 'kalian, 'kan, 'ludi, 'pamah, 'dipimpin, 'dari, 'partai, 'pdi, 'kalo, 'lalu, 'pgr, 'sesuatu, 'yg, 'beda, 'ya, 'di, 'pgr, 'ganti, 'u0002raab]	[not, 'hard, 'keeling, 'ya, 'ludi, 'dipimpin, 'partai, 'pdi, 'kalo, 'lalu, 'pgr, 'yg, 'beda, 'ya, 'di, 'pgr, 'ganti]
[bekas, 'full, 'prabowo]	[bekas, 'full, 'prabowo]
[hanura, 'masalah, 'yg, 'pny, 'tv, 'bisa, 'di, 'balikan, 'fakta, 'berta]	[hanura, 'yg, 'pny, 'tv, 'balikan, 'fakta, 'berta]
[prabowo]	[prabowo]
[insya, 'allah, 'pak, 'anis]	[insya, 'allah, 'anis]
[anis]	[anis]
[pak, 'anessss, 'yessss]	[anessss, 'yessss]
[pk, 'anis]	[pk, 'anis]
[anis, 'yess]	[anis, 'yess]
[anuhi, 'yg, 'menang]	[anuhi, 'yg, 'menang]
[pak, 'anis, 'tasid, 'baswedan, 'presiden]	[anis, 'tasid, 'baswedan, 'presiden]
[anis]	[anis]
[pan, 'pani, 'pani, 'ganjar, 'semakin, 'di, 'depan]	[pani, 'pani, 'pani, 'ganjar]

Gambar 4. 7 Tahap *Filtering*

f. Menggabungkan Kata Menjadi Kalimat Baru

Setelah teks diolah dengan cara sebelumnya, kata-kata yang tersisa dapat digabungkan kembali menjadi kalimat baru. Ini membantu mempertahankan konteks dan struktur kalimat dalam teks. *Script* yang digunakan adalah :

preprocessingwords = " ".join(filterwords)

Terakhir, hasil *filtering* kata-kata yang relevan digabungkan kembali menjadi sebuah kalimat baru menggunakan metode **join()**. Kalimat baru ini akan menjadi hasil akhir dari proses *preprocessing* teks. Adapun hasil data komentar yang telah melalui tahapan menggabungkan kata menjadi kalimat baru terlihat pada gambar di bawah ini.

Filtering	Join
[opk, 'prabowo', 'subianto', 'presiden', 'ku', '2024', 'yi']	bpk prabowo subianto presiden ku 2024 ni
[prabowo, 'presiden', 'gaspol', 'prabowo', 'kasih', 'kendor']	prabowo presiden gaspol prabowo kasih kendor
[prabowo, '']	prabowo
[apapun, 'kualifikasi', 'tetep', 'ganjar', 'presiden']	apapun kualifikasi tetep ganjar presiden
[prabowo, 'ir', '1']	prabowo 1
[prabowo, 'subianto', 'menang', 'pilpres', '2024']	prabowo subianto menang pilpres 2024
[prabowodwan, 'kamil', '']	prabowodwan kamil
[jilnas, 'alah', 'tdp', 'mu', 'pilih', 'prabowo', '']	jilnas alah tdp mu pilih prabowo
[prabowo, '2024']	prabowo 2024
[tangerang, 'banten', 'pilih', 'pa', 'anis']	tangerang banten pilih pa anis
[ganjar, '']	ganjar
[anis, 'maendo']	anis maendo
[prabowo, '']	prabowo
[masyarakat, 'butuh', 'perubahan']	masyarakat butuh perubahan
[pa, 'ganjar', '']	pa ganjar
[anis, 'pilih', 'rakya', 'no', 'wowo', 'no', 'uban']	anis pilih rakya no wowo no uban
[ganjar, 'pilihanku']	ganjar pilihanku
[prabowo, 'dukung', 'sengkumubur', 'bubar', '']	prabowo dukung sengkumubur bubar
[pilih, 'prabowo', 'males', 'golkar', 'pan', 'gara2', 'si', 'artis', 'pelakor', 'shahnaz', 'nyalon']	pilih prabowo males golkar pan gara2 si artis pelakor shahnaz nyalon
[klaten, 'jawa', 'dukung', 'anis', 'baswedan', 'cak', 'imin', 'capres', 'tawapres', '2024', '']	klaten jawa dukung anis baswedan cak imin capres tawapres 2024
[jgn, 'banteng', 'duras']	jgn banteng duras
[jateng, 'tetep', 'mlh', 'sp', 'prabowo']	jateng tetep mlh sp prabowo
[prabowo, '']	prabowo
[not, 'hard', 'feeling', 'ya', 'udh', 'dipimpin', 'partai', 'pdi', 'kalo', 'suh', 'pgr', 'yg', 'beda', 'ya', 'jd', 'pgr', 'gant', 'lu0001faad']	not hard feeling ya udh dipimpin partai pdi kalo suh pgr yg beda ya jd pgr gant
[bekasi, 'full', 'prabowo']	bekasi full prabowo
[hanura, 'yg', 'prnya', 'tv', 'balikan', 'takta', 'berita']	hanura yg prnya tv balikan takta berita
[prabowo, '']	prabowo
[insya, 'allah', 'anis']	insya allah anis
[anis, '']	anis
[anisssss, 'yesssss']	anisssss yesssss
[pk, 'anis']	pk anis
[anis, 'yesss']	anis yesss
[anish, 'yg', 'menang']	anish yg menang
[anis, 'tesid', 'baswedan', 'presiden']	anis tesid baswedan presiden
[anis, '']	anis
[ya, 'pan', 'pan', 'panjar']	pan pan pan panjar

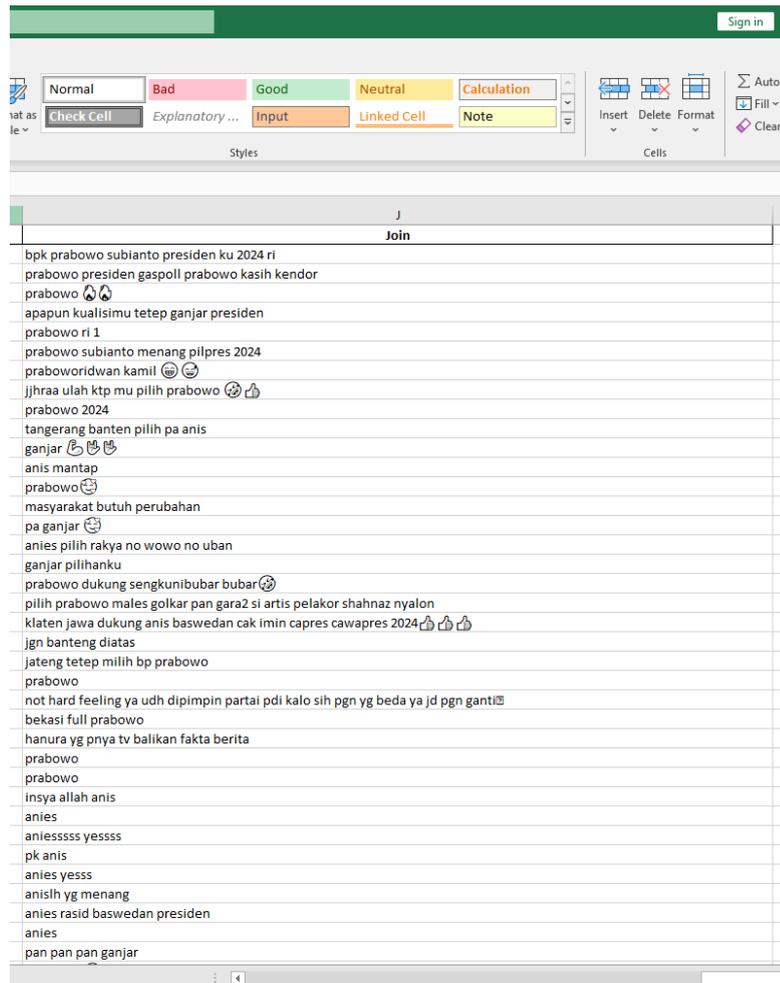
Gambar 4. 8 Tahap Menggabungkan Kata Menjadi Kalimat Baru

g. Ubah Data Lama Dengan yang Baru

Tahap terakhir adalah mengganti teks asli dengan teks yang telah diproses melalui langkah-langkah sebelumnya. *Script* yang digunakan adalah :

df.loc[i, "Text"] = preprocessingwords

Pada bagian ini, kita memiliki sebuah *dataframe* **df** yang berisi teks pada kolom **"Text"**. Tahap ini bertujuan untuk mengganti teks asli dalam kolom **"Text"** dengan hasil dari tahapan *preprocessing* yang telah dilakukan sebelumnya, yaitu **preprocessingwords**. Ini memastikan bahwa data yang diolah siap untuk digunakan dalam tahapan pengolahan data menggunakan *machine learning*. Adapun hasil data komentar yang telah melalui tahapan ubah data lama dengan yang baru terlihat pada gambar di bawah ini.



Gambar 4. 9 Tahap Ubah Data Lama Dengan yang Baru

4.3 Hasil Pengolahan Data Sentimen dengan *Machine Learning*

Setelah melalui serangkaian tahapan *preprocessing* teks yang meliputi *case folding*, penghapusan tanda baca, penghapusan *whitespace* yang tidak perlu, *tokenisasi*, *filtering*, dan penggabungan kembali kata-kata menjadi kalimat baru, teks komentar telah disiapkan untuk analisis selanjutnya. Langkah berikutnya dalam penelitian ini adalah menerapkan teknik *machine learning* untuk mengklasifikasikan komentar-komentar tersebut menjadi dua kategori: positif dan negatif. Dengan menggunakan data *training* sebanyak 1149 data yang telah disiapkan dan *epoch* 500, akan dibangun model *machine learning* yang dapat mengidentifikasi sentimen dari setiap komentar terhadap tiap pasangan capres dan cawapres.

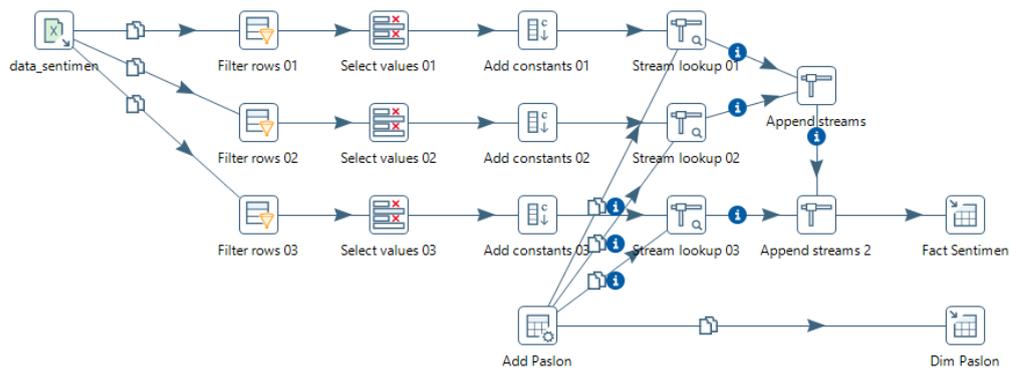
Penerapan *machine learning* akan memungkinkan kita untuk membuat model yang dapat memprediksi apakah suatu komentar bersifat positif atau negatif terhadap setiap capres dan cawapres. Hasil dari tahapan *machine learning* terlihat pada gambar di bawah ini.

Waktu	Text	User	Classification
18:04:22	bpk prabolo gan	bpk prabolo bpk prabolo bpk prabolo	p
14:20:44	prabowo i budiyant	prabowo i budiyant pak prabolo pak prabolo pak prabolo	p
16:03:11	prabowo i dafitbach	prabowo i dafitbach prabowo i prabowo i prabowo i	p
22:02:45	apapun ku qjaramala	sebesar aj sebesar aj sebesar aj [sebesar, 'apapun', apapun kualitasimu tetep ganjar presiden	p
14:21:28	prabowo i ukw_20	prabowo i	p
16:02:59	prabowo i user56753	pak prabolo pak prabolo pak prabolo [pak, 'pr'] prabowc prabowo subianto presiden gaspoll prabowo kasih kendur	p
15:48:15	prabowor @layArya	prabowor prabowor prabowor [prabowor] prabowor praboworidwan kamil	p
20:20:12	jhhraa ulaf mifta	@jhhraa_ jhhraa ulaf jhhraa ulaf [jhhraa, 'jhhraa', 'jhhraa ulah ktp mu pilih prabowo	p
15:41:42	prabowo i syahriela	prabowo i prabowo i prabowo i prabowo i prabowc prabowo 2024	p
13:49:44	tangerang ibu ade	siap saya, siap saya i siap saya i [siap, 'sa'] tangerantangeran banten pilih pa anis	p
00:11:16	ganjar	Dada rosit pak ganjar pak ganjar pak ganjar [pak, 'gar'] ganjar, 'ganjar	p
19:52:18	anis mant vfdy	anis mant anis mant anis mant [anis, 'mi'] anis, 'mi anis mantap	p
17:52:44	prabowo i lucky	prabowo i prabowo i prabowo i prabowo i prabowc prabowo	p
20:13:51	masyarakat tang subr	masyarakat masyarakat masyarakat [masyarakat] masyarakat masyarakat	p
19:22:45	pa ganjar jihan ni	pa ganjar pa ganjar pa ganjar [pa, 'gan'] pa ganjar	p
03:58:05	anies pilih suradja	anies di pianies di pianies di pi [anies, 'd'] anies, 'd anies pilih rakyat no wowo no uban	p
01:36:59	ganjar pili joniindra	pak ganjar pak ganjar pak ganjar pak ganjar [gar, 'ganjar', 'ganjar pilihhanku	p
14:33:40	prabowo i SWEATER	prabowo i prabowo i prabowo i prabowc prabowo dukung sengkinububar bubar	p
12:58:39	pilih prabau mom	mau pilih mau pilih [mau, 'pi'] pilih, 'pr pilih prabowo males golkar pan gara2 si artis pelakor shahnaz nyalon	p
18:06:24	klaten jav Bejo Alex	aku klater aku klater aku klater [aku, 'kla'] klaten, 'klaten jawa dukung anis baswedan cak imin capres cawapres 2024	p
21:45:36	ign bante daniel	ign ada la ign ada la [ign, 'ada'] ign, 'bi ign banteng diatas	n
22:45:35	jateng tet wim	jateng tet jateng tet jateng tet [jateng, 'jateng', 'jateng tetep milih bp prabowo	p
21:17:00	prabowo Arjun24	prabowo prabowo prabowo [prabowc] prabowc prabowo	p
16:41:51	not hard f Povidone	not hard f not hard f not hard f [not, 'har'] not, 'har not hard feeling ya udi dipimpin partai pidi kalo sih pgn yg beda ya jd pgn ganti	p
21:13:40	bekasi ful Bob_808	bekasi ful bekasi ful bekasi ful [bekasi, 'bekasi', 'bekasi full prabowo	p
16:56:00	hanura yg Blcksky7	hanura mi hanura mi hanura mi [hanura, 'hanura', hanura yg pnya tv balikan fakta berita	n
18:51:06	prabowo yudex @h	prabowo prabowo [prabowc] prabowc prabowo	p
22:32:53	prabowo rudhy263	prabowo prabowo prabowo [prabowc] prabowc prabowo	p
17:13:05	innya allat panwito45	innya allat innya allat innya allat [innya, 'a'] innya, 'a innya allah anis	p
17:17:34	anies AnakStam	anies anies [anies, 'anies'] anies	p
13:41:03	aniesssss asari9283	pak aniesi pak aniesi pak aniesi [pak, 'ani'] aniesssss aniesssss yessss	p
19:27:20	pk anis Fathorrah	pk anis pk anis [pk, 'anis'] anis pk anis	p
15:05:34	anies yess Ahmad R	anies yess anies yess anies yess [anies, 'y'] anies, 'y anies yess	p
11:38:36	anisih yg rucuu sumi	anisih yg r anisih yg r anisih yg r [anisih, 'y'] anisih, 'y anisih yg menang	p
18:03:15	anies rasi h, nur has	pak anies pak anies [pak, 'ani'] anies, 'ri anies rasid baswedan presiden	p
17:40:03	anies Alyasri393	anies anies [anies, 'anies'] anies	p
14:30:48	pan pan pi Lot Tik	pan pan p pan pan p pan pan p [pan, 'pai'] pan, 'pa pan pan pan ganjar	p

Gambar 4. 10 Hasil Klasifikasi Data Sentimen dengan *Machine Learning*

4.4 Analisis Data dengan Pentaho

Setelah melakukan tahap *machine learning* untuk mengklasifikasikan komentar-komentar menjadi kategori positif dan negatif terhadap setiap capres, tahap berikutnya dalam penelitian ini, yaitu integrasi data menggunakan *platform* Pentaho. Integrasi data adalah proses menggabungkan data dari berbagai sumber, termasuk data hasil analisis sentimen yang telah diperoleh dari tahap sebelumnya, untuk memungkinkan analisis yang lebih mendalam. Adapun skema data analisis menggunakan pentaho terlihat pada gambar di bawah ini.



Gambar 4. 11 Skema Analisis Data Menggunakan *Pentaho Data Integration*

Penjelasan detail dari skema analisis data sentimen menggunakan *Pentaho Data Integration* adalah sebagai berikut :

a. Data Sentimen *Excel*

Langkah pertama adalah mengimpor data sentimen dari file *Excel* ke dalam PDI. File *Excel* ini mungkin berisi teks-teks yang akan dianalisis sentimennya, serta atribut-atribut tambahan seperti tanggal, pengguna, dan komentar.

b. *Filter Rows*

Setelah data diimpor, langkah berikutnya adalah menerapkan filter baris untuk membersihkan data. Filter baris digunakan untuk mengekstrak subset data yang memenuhi kriteria tertentu. Misalnya dapat menggunakan filter baris untuk menghapus baris yang kosong atau menghilangkan entri yang tidak relevan. Pada tahapan ini, *filter rows* yang digunakan ada 3 sesuai dengan calon capres 2024. Adapun kondisi yang digunakan pada *select rows* 01 adalah :

01 = [positif]

OR

01 = [positif]

Catatan : lakukan hal yang sama pada *select rows* 02 dan 03 (rubah angka sesuai dengan 02 dan 03).

c. *Select Value*

Langkah ini digunakan untuk memilih nilai-nilai tertentu dari kolom-kolom dalam dataset. Ini bisa berguna untuk mengambil kolom-kolom tertentu yang akan digunakan dalam analisis sentimen, atau untuk mengekstrak informasi

penting dari data lainnya. Pada *fieldname* “**Tanggal**” *rename to* “**:tanggal**”. Pada *fieldname* “**01**” *rename to* “**:sentimen**”.

d. *Add Constant*

Tahap ini dapat digunakan untuk menambahkan nilai konstan ke dalam dataset. Ini bisa berguna jika Anda perlu menambahkan kolom baru yang memiliki nilai yang sama untuk setiap baris dalam dataset, seperti label sentimen awal (misalnya, "positif", "negatif"). Pada *add constant*, kolom *name* diisi dengan **no_paslon**, *type string*, *value 01*, dan *set empty string* diisi dengan N.

e. *Stream Lookup*

Langkah ini mungkin digunakan untuk mengaitkan data dari satu sumber dengan data dari sumber lainnya berdasarkan kunci tertentu. Misalnya dapat menggunakan *stream lookup* untuk mengaitkan data sentimen dengan data lain, seperti data tentang pasangan calon presiden, berdasarkan kunci yang sama (misalnya, nama calon presiden pada nomor paslon). Pada *stream lookup*, kolom *Field* diisi dengan **no_paslon** dan *LookupField* diisi dengan **no_paslon**.

f. *Append Stream*

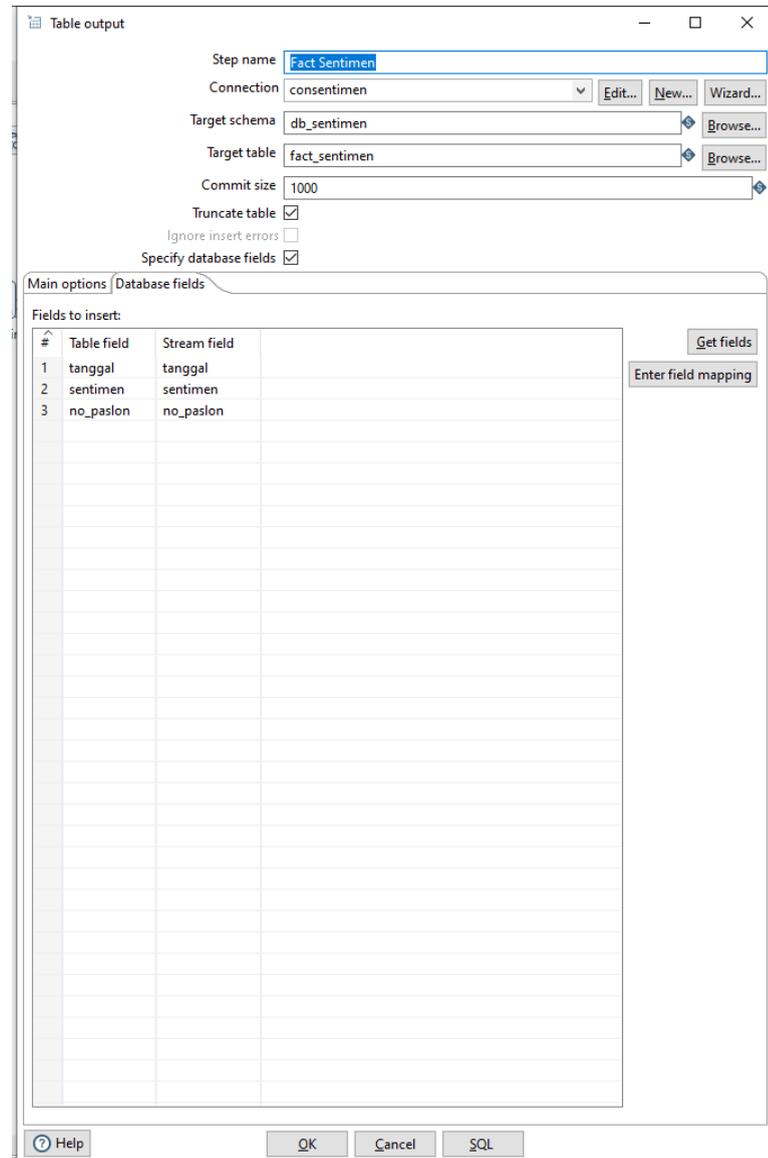
Langkah ini digunakan untuk menggabungkan atau menambahkan data dari dua sumber yang berbeda menjadi satu. Ini berguna jika Anda memiliki dua set data yang ingin Anda gabungkan menjadi satu, seperti data sentimen dan data tentang pasangan calon presiden. Pada *append stream*, *Head hop* diisi dengan **stream lookup 01** dan *tail hop* diisi dengan **stream lookup 02**. Begitu pula dengan *append streams 2*, *Head hop* diisi dengan **append stream** dan *tail hop* diisi dengan **stream lookup 03**.

g. *Add Paslon*

Langkah ini digunakan untuk menambahkan informasi tentang pasangan calon presiden ke dalam dataset termasuk nama dan nomor paslon. Pada kolom *name* diisi dengan **no_paslon dan nama_paslon**, pada kolom *type* diisi dengan **string**, pada kolom *length* diisi dengan **2 dan 100**, pada kolom *set empty string* diisi dengan N.

h. *Fact Sentimen*

Tahap ini digunakan untuk membuat tabel fakta yang berisi informasi tentang sentimen dari teks-teks yang dianalisis. Isi dari *fact sentimen* terlihat pada gambar di bawah ini.

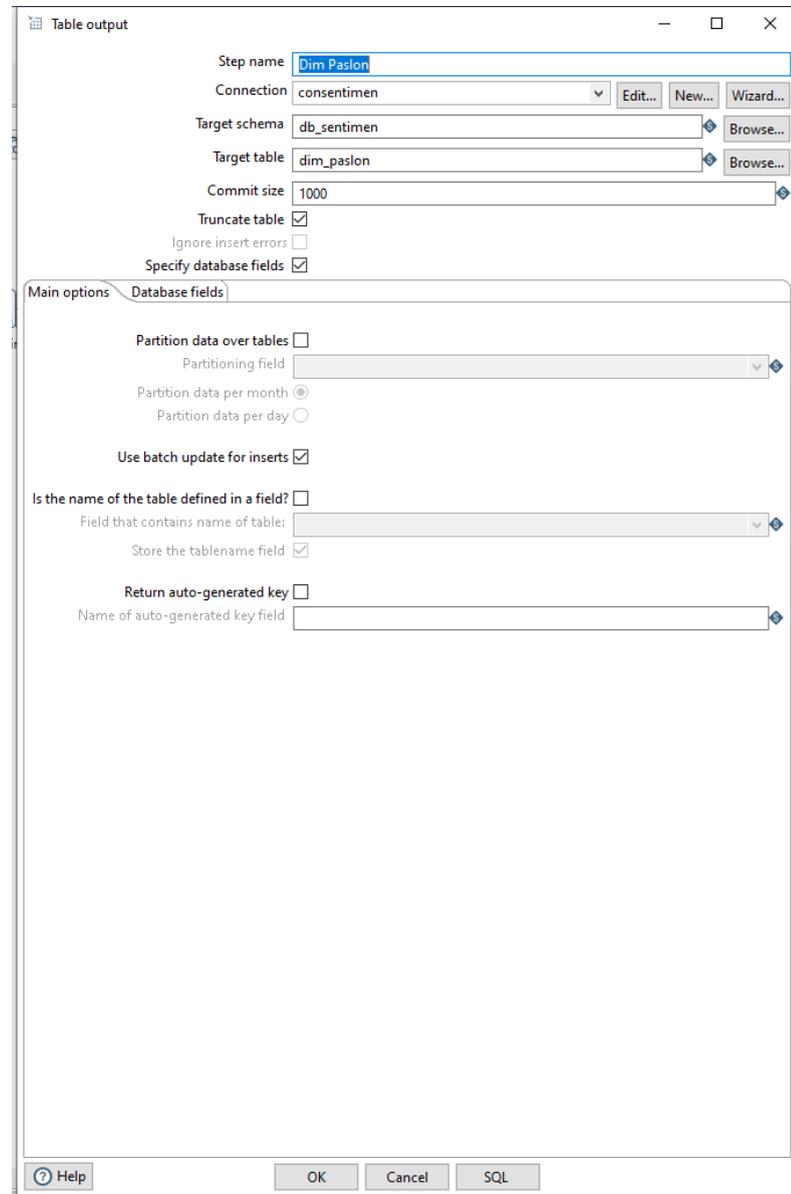


Gambar 4. 12 Isi dari *Fact Sentimen*

i. Dim Paslon

Langkah terakhir adalah membuat dimensi pasangan calon presiden. Dimensi ini berisi informasi tentang pasangan calon presiden yang relevan dengan analisis sentimen, seperti nama dan nomor paslon yang dapat digunakan untuk

menganalisis hubungan antara sentimen dan paslon. Isi dari dim paslon terlihat pada gambar di bawah ini.



Gambar 4. 13 Isi Dim Paslon

Gambar di atas dijelaskan bahwa hasil dari skema *Pentaho Data Integration* memiliki 2 tabel data, yaitu tabel fakta dan tabel dimensi. Tabel data dan dimensi dapat dilihat pada aplikasi **Xampp** seperti gambar di bawah ini. Tabel fakta berupa sentimen, sedangkan tabel dimensi berupa paslon (capres/cawapres).



The image shows two database table schemas. The first is 'fact_sentimen' with columns: id_sentimen (int(11)), tanggal (date), no_paslon (varchar(2)), and sentimen (varchar(1)). The second is 'dim_paslon' with columns: no_paslon (varchar(2)) and nama_paslon (varchar(100)).

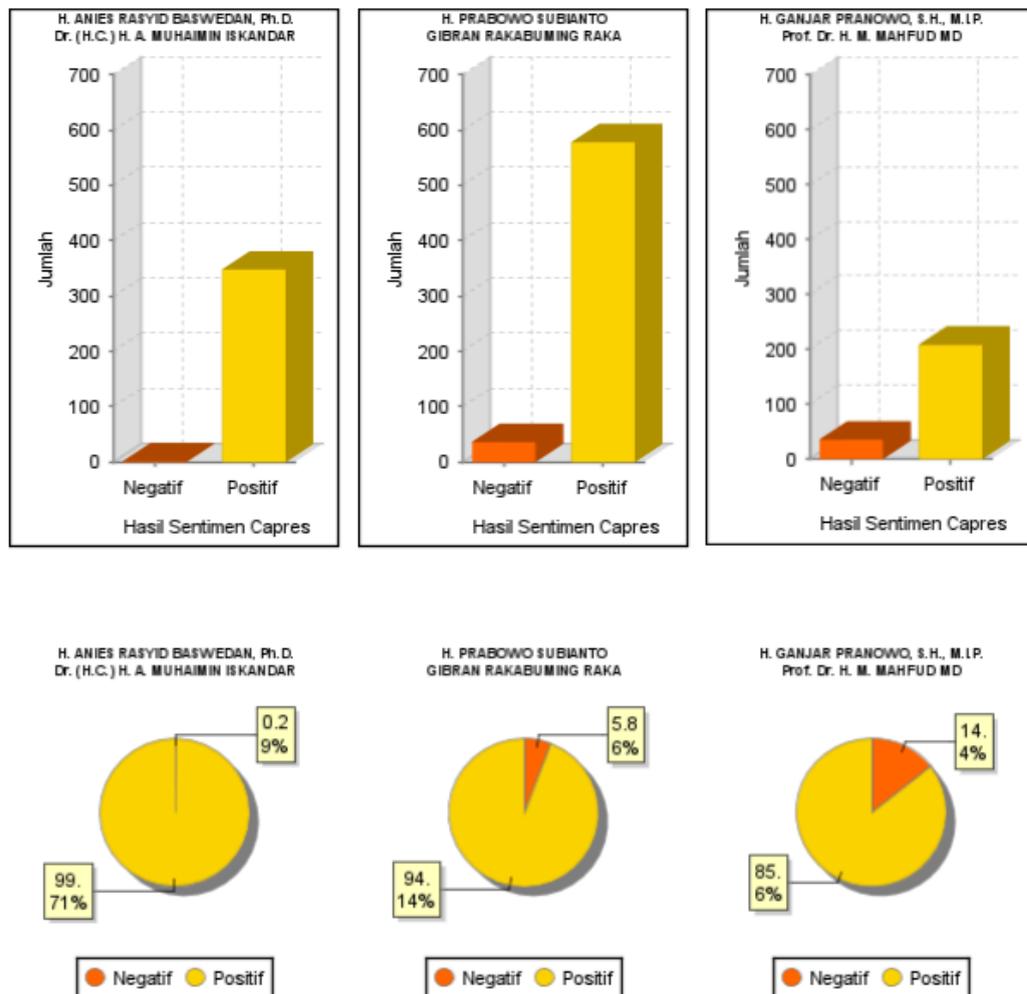
Table Name	Column Name	Data Type
db_sentimen fact_sentimen	id_sentimen	int(11)
	tanggal	date
	no_paslon	varchar(2)
	sentimen	varchar(1)
db_sentimen dim_paslon	no_paslon	varchar(2)
	nama_paslon	varchar(100)

Gambar 4. 14 Tabel Fakta dan Tabel Dimensi

4.5 Hasil Sentimen

Setelah menyelesaikan proses integrasi data menggunakan *Pentaho Data Integration* dan memahami skema data yang dihasilkan, tahap berikutnya yaitu membuat grafik capres dan cawapres hasil analisis sentimen dari tabel fakta menggunakan *Pentaho Designer*. Grafik hasil analisis sentimen capres dan cawapres 2024 yang diambil dari komentar TikTok terlihat pada gambar di bawah ini. Jumlah sentimen pasangan

HASIL ANALISIS SENTIMEN CAPRES 2024



Gambar 4. 15 Grafik Hasil Analisis Sentimen Berdasarkan Komentar TikTok

4.6 Kelebihan dan Kekurangan

Analisis sentimen terhadap komentar TikTok mengenai calon presiden (capres) dan calon wakil presiden (cawapres) 2024 memiliki sejumlah kelebihan dan kekurangan, yaitu :

a. Kelebihan

Kelebihan dari analisis sentimen terhadap komentar TikTok mengenai calon presiden (capres) dan calon wakil presiden (cawapres) 2024 adalah :

1. TikTok adalah platform yang populer dan memiliki jutaan pengguna aktif. Oleh karena itu, analisis sentimen berdasarkan komentar TikTok dapat memberikan akses ke sumber data yang beragam.
2. Komentar-komentar yang diposting di TikTok cenderung memberikan pandangan yang langsung dan aktual dari masyarakat. Ini memungkinkan analisis sentimen untuk memberikan wawasan yang cepat dan *real-time* tentang persepsi dan opini terbaru terkait dengan calon presiden dan wakil presiden.

b. Kekurangan

Kekurangan dari analisis sentimen terhadap komentar TikTok mengenai calon presiden (capres) dan calon wakil presiden (cawapres) 2024 adalah komentar di TikTok sering menggunakan bahasa informal, singkatan, atau slang yang sulit dipahami oleh algoritma analisis sentimen. Hal ini dapat menghasilkan interpretasi yang kurang akurat atau bahkan salah terhadap sentimen yang diekspresikan.