

BAB III

METODE PENELITIAN

3.1 Analisa Kebutuhan

Analisa kebutuhan merupakan proses penting dalam pendataan dan identifikasi kebutuhan yang diperlukan pada suatu proyek atau penelitian. Analisa kebutuhan bertujuan untuk memastikan penelitian dapat berjalan lancar sesuai dengan tujuan yang diinginkan. Melalui Analisa kebutuhan yang menyeluruh, identifikasi kebutuhan menjadi jelas dan mendetail sehingga dapat memfasilitasi pelaksanaan penelitian yang efektif.

3.1.1 Kebutuhan Sistem

Kebutuhan sistem pada analisis sentimen ini akan dibagi menjadi 2, yaitu kebutuhan fungsional dan kebutuhan non-fungsional. Kedua kebutuhan ini diperlukan demi berjalannya proses penelitian yang baik. Pengertian dari kebutuhan fungsional adalah kebutuhan ini membuat sistem mampu bereaksi pada situasi yang tepat. Salah satu alasan yang tepat ialah kebutuhan fungsional ini harus menggambarkan secara jelas mengenai kegunaan sistem, seperti proses apa yang harus dijalankan (Liliana, 2021). Pada penelitian ini pengambilan data pada komentar akan menggunakan Instant data scraper dengan jumlah komentar yang telah terkumpul sebanyak 1020 komentar, hasil komentar yang didapat mulai dari bulan 1 Agustus 2023 hingga 5 Agustus 2024.

Kebutuhan *non-functional* adalah suatu kebutuhan yang ditambahkan sebagai hal-hal yang diperlukan untuk membantu proses sistem sehingga fokusnya pada batasan layanan. Dalam kebutuhan *non-functional* ini akan terdapat dua bagian yaitu perangkat lunak (*Software*) dan perangkat keras (*Hardware*) yang didasarkan pada kebutuhan *non-functional* pada analisis sentimen :

- a. Perangkat Keras (*Hardware*):

Tabel 3.1 Perangkat Keras

Hardware	Keterangan
<i>Processor</i>	Intel Core i3
<i>Memory</i>	4 GB
SSD	238,47 GB

- b. Perangkat Lunak (*Software*):

Tabel 3.2 Perangkat Lunak

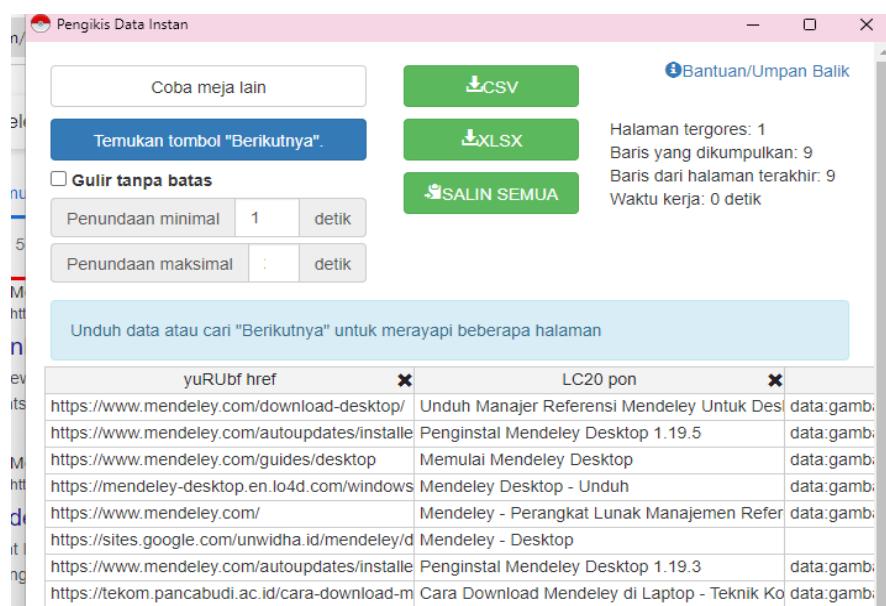
Software	Keterangan
Sistem Operasi	Windows 11
<i>Web Browser</i>	Microsoft Edge
Instant Data Scraper	Pengaksesan fitur data
Microsoft Excel	2019
Google Colab	Menjalankan Python

3.2 Perancangan Penelitian

Pada proses perancangan penelitian nantinya akan dibuat proses perancangan menjadi pembagian tertentu yaitu *scraping*, *cleaning*, *labelling*, *case folding*, *filtering/stopwords*, *tokenizing*, *stemming*, dan *term frequency*.

3.2.1 Data Set

Data ini merupakan suatu data yang ada di publik dimana pada penelitian ini akan diambil dari komentar Youtube, untuk mendapatkan data tersebut tekniknya dinamakan *scraping*.

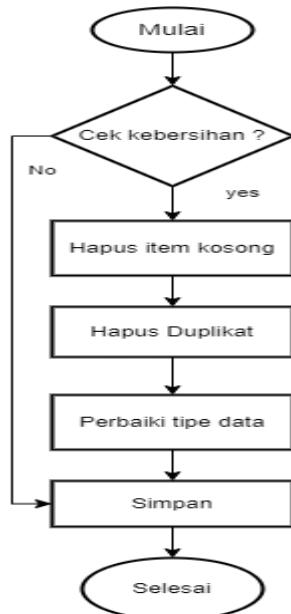


Gambar 3.1 Data Set

Diawali dengan proses pengambilan data dengan menggunakan Instant Data Scraper. Dikarenakan sumber dari proses *scraping* data ini dari youtube maka akan lebih mudah untuk mendapatkan data komentar melalui fasilitas layanan tersebut.

3.2.2 Cleaning

Pada bagian *cleaning* dilakukan pembersihan dengan cara menghilangkan karakter dan tanda baca yang tidak perlu, berikut dibawah ini merupakan langkah-langkahnya.



Gambar 3.2 Tahap Cleaning

Melihat dari langkah *cleaning*, perlu diketahui dalam melihat data yang telah didapatkan, apakah data perlu dilakukan *cleaning* atau data sudah bersih maka bisa di check terlebih dahulu.

```

✓ 6s  ⏴ pip install Sastrawi
  ↗ Collecting Sastrawi
    Downloading Sastrawi-1.0.1-py2.py3-none-any.whl (209 kB)
    ━━━━━━━━━━━━━━━━ 209.7/209.7 kB 2.0 MB/s eta 0:00:00
  Installing collected packages: Sastrawi
  Successfully installed Sastrawi-1.0.1

```

punkt untuk menghapus simbol dan kata kata yang tidak perlu

```

✓ 4s [12] import nltk
      nltk.download('punkt');

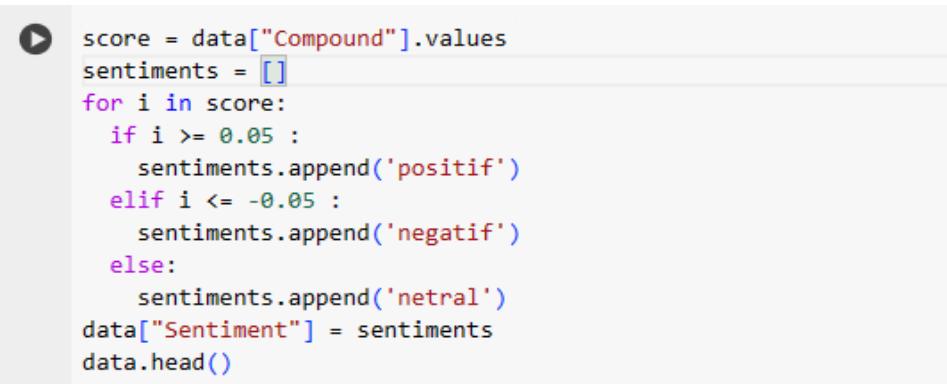
```

Gambar 3.3 Code Python Cleaning

Satrawi merupakan *library* python untuk pemrosesan bahasa alami dalam konteks bahasa indonesia sehingga pengolahan data dan analisis teks Bahasa Indonesia lebih akurat. Jika telah diketahui data tersebut harus dibersihkan maka lakukan penghapusan item kosong, hapus duplikat kata, emoji yang tidak diperlukan, tanda seru dan perbaiki tipe data. Punkt biasa digunakan untuk pemecah teks, yang tugasnya mengidentifikasi tanda titik, tanya, dan seru.

3.2.3 *Labelling*

Labelling dilakukan dengan cara melakukan pembedaan pada hasil komentar menjadi positif, negatif dan netral, sebagai contoh data akan dibagi menjadi tiga secara acak berdasarkan proses yang tadi sudah dilalui dengan Instant Data Scraper.



```

score = data["Compound"].values
sentiments = []
for i in score:
    if i >= 0.05 :
        sentiments.append('positif')
    elif i <= -0.05 :
        sentiments.append('negatif')
    else:
        sentiments.append('netral')
data["Sentiment"] = sentiments
data.head()

```

Gambar 3.4 *Code Python Labelling*

Code ini menilai sentimen dari setiap nilai dalam kolom "*Compound*" dan menambahkan label sentimen ("positif", "negatif", atau "netral") ke *Dataframe* data pada kolom baru yang disebut "*Sentiment*". Berikut ini adalah contoh data yang telah berlabel :

Tabel 3.3 Hasil *Labelling*

No.	Komentar	Sentimen
1.	Saya setuju dengan pendapat ini karena <i>supplay</i> dan demennya penghargaan profesi guru sangat rendah di negeri ini.	Positif
2.	Semoga yang disini bisa jadi guru sukses yang lebih kreatif lagi supaya gaji mumpuni untuk kebutuhan hidup.	Positif
3.	Salah satu hal terpenting untuk menjadi seorang guru adalah memiliki gairah terhadap pendidikan dan keinginan yang tulus untuk membantu siswa belajar dan berkembang.	Positif
4.	gua setuju jika semua orang bisa jadi Guru, tetapi profesi guru bukan sekedar mengajar tetapi bagaimana mendidik murid agar lebih baik dalam memahami hidup. sekedar mengajar semua bisa kok tetapi tak mampu bagaimana mengubah <i>mindset</i> pola pikir murid.	Netral

5.	Saya pikir semua pekerjaan membutuhkan keahlian,dan mungkin besaran gaji yg diterima pekerja itu sebanding dengan keterampilan mengatasi seberapa sulit pekerjaan itu dibebankan.	Netral
6.	katanya murid mesti belajar, gurunya sendiri kagak mau belajar teknologi (banyak juga yang gaptek).	Negatif
7.	Banyak guru yg ke sarkas ngga terima kenyataan bahwa jadi guru bukan hal yang susah semua orang bisa jadi guru kalo dlu pemikiran kita di bangun untuk meyakini kalo guru adalah pahlawan tanpa jasa awokawok padahal cuma berlaku untuk guru SD yang ngajar dari kita belum tau apa apa ngajarin baca ngitung dan lain lain, komentar ini termasuk positif, netral atau negatif.	Negatif
8.	Gak salah sih bang tapi ya gak sepenuhnya benar pendapat lho Menurut gw guru di gaji rendah itu karena standarisasi gaji berdasarkan jumlah jam mengajar bukan berdasarkan kualifikasi. Terus rendahnya gaji guru honorer terjadi karena proses perekrutan yg salah karena direkrut	Negatif

	bukan berdasarkan kebutuhan, melainkan jalur orang dalam. Dan sertifikasi hanya dilakukan satu kali seumur hidup bukan periodik, sehingga yg sudah serti biasanya kerjanya jadi menggampangkan. Jadi dia yg udah serti bisa kipas adem ayem tanpa beban Jadi kayak ada dikotomi sendiri dalam karir keguruan Miris tapi mau gimana lagi memang sistem dan kebijakan masih belum bagus amat. Ditambah moralitas yg bobrok. Gue setuju bang semua orang bisa mengajar. Tapi tidak semua orang bisa mendidik. Karena mendidik itu bukan sekedar menyampaikan teori kelimuan, tapi juga karakter, adap, sopan santun dan masih banyak lagi.	
9.	Yang saya sayangkan saya sendiri lulus dari FKIP adalah teorinya 95% tetapi praktik mengajarnya dalam artian seluk beluk kependidikan utk memanage kelas dan siswa itu hanya 5%. Sehingga mrk dibiarkan utk apply ngajar sendiri di luar. Pembahasan ttg mengajarnya yg secara terspesialisasi itu tidak ada di bangku universitas. Mahasiswa juga	Negatif

	tidak diajarkan untuk bagaimana bisa percaya diri taunya cuma nyuruh-nyuruh aja. Dalam pikiran dosennya taunya mhs udah bisa aja. <i>Value</i> pendidikan di sini jadinya berkurang. Dosenya juga saya katakan punya kecerdasan emosi yg rendah karna tiba tiba suka marah marah aja di kelas mahasiswa jadi pelampiasan emosi dosennya yg mana menurut hemat saya pendidik seharusnya tidak seperti itu. Alangkah baiknya jika pendidik harus teruji IQ dan EQ nya tidak hny melalui tes tapi di praktiknya juga. Betul betul diajarkan secara spesialisasi seperti dokter lah harus terkhususkan utk mencetak pengajar dan pendidik yg berkualitas.	
10.	Gelar S.Pd sebagai Sarjana pendidikan malah menjadi Sarjana Penuh Derita.	Negatif

3.2.4 *Case Folding*

Selanjutnya akan dilakukan bagian *case folding*, setelah melalui proses *cleaning*.



Gambar 3.6 Tahap *Case Folding*

Case folding yang akan digunakan untuk mengubah atau dapat dikatakan membuat semua huruf menjadi huruf kecil semua sehingga huruf lebih konsisten.

```

import pandas as pd

# Membaca data dari file Excel
data = pd.read_excel("youtubeSkripsiGuru.xlsx")

# Mengonversi teks menjadi huruf kecil pada kolom 'style-scope 2'
data['style-scope 2'] = data['style-scope 2'].apply(lambda x: x.lower() if isinstance(x, str) else x)

# Menampilkan data setelah pengonversian
print(data.head())
  
```

Gambar 3.5 *Code Python Case Folding*

Proses *case folding* melibatkan konversi teks menjadi huruf kecil khusus pada kolom yang disebut *style-scope 2*, kolom tersebut berisi data yang telah mengalami proses *cleaning*. Data ini didapatkan dari data excel bernama “YoutubeSkripsiGuru”. Hasil proses *case folding* dapat ditampilkan untuk dilihat sebagai berikut :

Tabel 3.4 Hasil Case Folding

No.	Data hasil Cleaning	Data yang telah di Case Folding
1.	Gaji guru di Indonesia dibanding pns yang lain sudah punya gaji yang lebih bagus buktinya semua guru sudah punya mobil dan rumah yang layak kalau mereka bilang cuma punya sepeda ontel itu bohong besar karena kemaruk mereka sebab disamping dia punya gaji juga ada remunerasi yang besarnya sama dengan gajinya. yang perlu diperhatikan adalah penempatannya. banyak sekolah kebanyakan guru sementara didaerah lainnya kekurangan guru. mestinya mereka mau ditempatkan diseluruh wilayah indonesia sebagaimana perjanjian semula, nyatanya banyak yang tidak mau dan hanya ingin dikota besar	gaji guru di indonesia dibanding pns yang lain sudah punya gaji yang lebih bagus buktinya semua guru sudah punya mobil dan rumah yang layak kalau mereka bilang cuma punya sepeda ontel itu bohong besar karena kemaruk mereka sebab disamping dia punya gaji juga ada remunerasi yang besarnya sama dengan gajinya. yang perlu diperhatikan adalah penempatannya. banyak sekolah kebanyakan guru sementara didaerah lainnya kekurangan guru. mestinya mereka mau ditempatkan diseluruh wilayah indonesia sebagaimana perjanjian semula, nyatanya banyak yang tidak mau dan hanya ingin dikota besar
2.	Salah satu hal terpenting untuk menjadi seorang Guru adalah	salah satu hal terpenting untuk menjadi seorang guru adalah memiliki gairah

	memiliki gairah terhadap pendidikan dan keinginan yang tulus untuk membantu Siswa belajar dan berkembang	terhadap pendidikan dan keinginan yang tulus untuk membantu siswa belajar dan berkembang
3.	Tolong bang bahas juga berapa jumlah angka kemiskinan Indonesia dilihat dari tingkat Pendidikan	tolong bang bahas juga berapa jumlah angka kemiskinan indonesia dilihat dari tingkat pendidikan
4.	Bukan profesi Guru nya tpi lebih ke sistem di negara ini yg aneh	bukan profesi guru nya tpi lebih ke sistem di negara ini yg aneh
5.	Guru tidak hanya mengajar tapi juga mendidik semua orang mungkin bisa mengajar tapi hanya Guru yang bisa mendidik	guru tidak hanya mengajar tapi juga mendidik semua orang mungkin bisa mengajar tapi hanya guru yang bisa mendidik
6.	<i>Hello Bang Evan menyimak yaaa terima kasih banyak sudah menyajikan konten ini sedemikian rupa yg sangat sistematis dan runut berikut dg referensi-referensinya. sukses selalu buat tim satu persen</i>	<i>hello bang evan menyimak yaaa terima kasih banyak sudah menyajikan konten ini sedemikian rupa yg sangat sistematis dan runut berikut dg referensi-referensinya. sukses selalu buat tim satu persen</i>

7.	Sebenarnya cuma satu caranya perubahan keseluruhan sistem pendidikan di Indonesia dengan begitu akan merubah semuanya perubahan dengan lebih baik dan lebih maju tentunya	sebenarnya cuma satu caranya perubahan keseluruhan sistem pendidikan di indonesia dengan begitu akan merubah semuanya perubahan dengan lebih baik dan lebih maju tentunya
8.	Sangat Rill untuk <i>recruit</i> Guru honorer khususnya di kampung sangat mudah klo ada orang dalam dulu pas saya SD tukang jaga sekolah dan jualan di kantin bisa jadi guru olahraga yang rada <i>toxic</i> dan ngomong kasar menjadikan murid ngikutin ngomong kasar	sangat rill untuk <i>recruit</i> guru honorer khususnya di kampung sangat mudah klo ada orang dalam dulu pas saya sd tukang jaga sekolah dan jualan di kantin bisa jadi guru olahraga yang rada <i>toxic</i> dan ngomong kasar menjadikan murid ngikutin ngomong kasar
9.	Mungkin karna menggurui itu di benci guru yang benar bukan menggurui tapi mengerti keinginan murid dan tau apa yang harus dilakukan guru ke murid	mungkin karna menggurui itu di benci guru yang benar bukan menggurui tapi mengerti keinginan murid dan tau apa yang harus dilakukan guru ke murid
10.	Saya pikir semua pekerjaan membutuhkan keahlian dan	saya pikir semua pekerjaan membutuhkan keahlian dan mungkin

	mungkin besaran gaji yang diterima pekerja itu sebanding dengan keterampilan mengatasi seberapa sulit pekerjaan itu dibebankan	besaran gaji yang diterima pekerja itu sebanding dengan keterampilan mengatasi seberapa sulit pekerjaan itu di bebankan
--	--	---

3.2.5 *Tokenizing*

Tahap *tokenizing* adalah tahap pemisahan kalimat komentar yang panjang menjadi kata Tunggal, dimulai dari kata awal sampai akhir dari suatu kalimat. Dibawah ini merupakan langkah-langkah *tokenizing*.



Gambar 3.7 Tahap *Tokenizing*

Tokenizing berguna untuk membagi setiap kata sehingga akan terbagi menjadi unit-unit kecil dan merupakan salah satu proses penting dalam penelitian ini. Pembagian ini dilakukan setelah *case folding*, kata-kata yang dipisahkan telah dipilih terlebih dahulu.

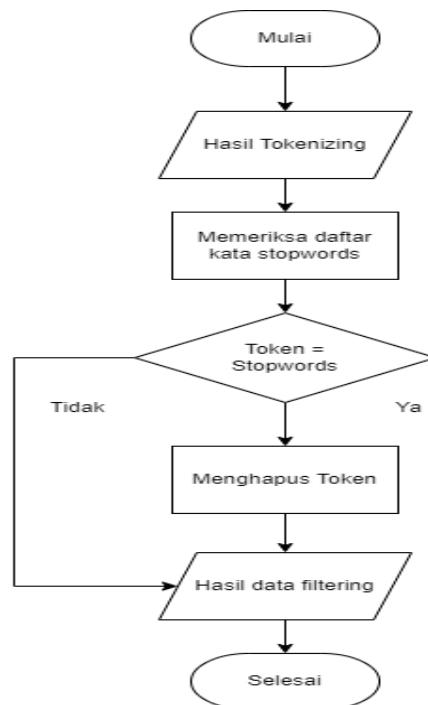
Tabel 3.5 Hasil *Tokenizing*

No.	Data yang telah di <i>Tokenizing</i>
1.	[‘gaji’, ‘guru’, ‘di’, ‘indonesia’, ‘dibanding’, ‘pns’, ‘yang’, ‘lain’, ‘sudah’, ‘punya’, ‘gaji’, ‘yang’, ‘lebih’, ‘bagus’, ‘buktinya’, ‘semua’, ‘guru’, ‘sudah’, ‘punya’, ‘mobil’, ‘dan’, ‘rumah’, ‘yang’, ‘layak’, ‘kalau’, ‘mereka’, ‘bilang’, ‘cuma’, ‘punya’, ‘sepeda’, ‘ontel’, ‘itu’, ‘bohong’, ‘besar’, ‘karena’, ‘kemaruk’, ‘mereka’, ‘sebab’, ‘disamping’, ‘dia’, ‘punya’, ‘gaji’, ‘juga’, ‘ada’, ‘remunerasi’, ‘yang’, ‘besarnya’, ‘sama’, ‘dengan’, ‘gajinya’, ‘yang’, ‘perlu’, ‘diperhatikan’, ‘adalah’, ‘penempatannya’, ‘banyak’, ‘sekolah’, ‘kebanyakan’, ‘guru’, ‘sementara’, ‘didaerah’, ‘lainnya’, ‘kekurangan’, ‘guru’, ‘mestinya’, ‘mereka’, ‘mau’, ‘ditempatkan’, ‘diseluruh’, ‘wilayah’, ‘indonesia’, ‘sebagaimana’, ‘perjanjian’, ‘semula’, ‘nyatanya’, ‘banyak’, ‘yang’, ‘tidak’, ‘mau’, ‘dan’, ‘hanya’, ‘ingin’, ‘dikota’, ‘besar’]
2.	[‘salah’, ‘satu’, ‘hal’, ‘terpenting’, ‘untuk’, ‘menjadi’, ‘seorang’, ‘guru’, ‘adalah’, ‘memiliki’, ‘gairah’, ‘terhadap’, ‘pendidikan’, ‘dan’, ‘keinginan’, ‘yang’, ‘tulus’, ‘untuk’, ‘membantu’, ‘siswa’, ‘belajar’, ‘dan’, ‘berkembang’]
3.	[‘tolong’, ‘bang’, ‘bahas’, ‘juga’, ‘berapa’, ‘jumlah’, ‘angka’, ‘kemiskinan’, ‘indonesia’, ‘dilihat’, ‘dari’, ‘tingkat’, ‘pendidikan’]

4.	['bukan', 'profesi', 'guru', 'nya', 'tpi', 'lebih', 'ke', 'system', 'di', 'negara', 'ini', 'yg', 'aneh']
5.	['guru', 'tidak', 'hanya', 'mengajar', 'tapi', 'juga', 'mendidik', 'semua', 'orang', 'mungkin', 'bisa', 'mengajar', 'tapi', 'hanya', 'guru', 'yang', 'bisa', 'mendidik']
6.	['hello', 'bang', 'evan', 'menyimakk', 'yaaa', 'terima', 'kasih', 'banyak', 'sudah', 'menyajikan', 'konten', 'ini', 'sedemikian', 'rupa', 'yg', 'sangat', 'sistematis', 'dan', 'runut', 'berikut', 'dg', 'referensi-referensinya', 'sukses', 'selalu', 'buat', 'tim', 'satu', 'persen']
7.	['perubahan', 'sistim', 'pendidikan', 'indonesia', 'merubah', 'perubahan', 'maju']
8.	['sangat', 'rill', 'untuk', 'recruit', 'guru', 'honorer', 'khususnya', 'di', 'kampung', 'sangat', 'mudah', 'klo', 'ada', 'orang', 'dalam', 'dulu', 'pas', 'saya', 'sd', 'tukang', 'jaga', 'sekolah', 'dan', 'jualan', 'di', 'kantin', 'bisa', 'jadi', 'guru', 'olahraga', 'yang', 'rada', 'toxic', 'dan', 'ngomong', 'kasar', 'menjadikan', 'murid', 'ngikutin', 'ngomong', 'kasar']
9.	['mungkin', 'karna', 'menggurui', 'itu', 'di', 'benci', 'guru', 'yg', 'bener', 'bukan', 'menggurui', 'tapi', 'mengerti', 'keinginan', 'murid', 'dan', 'tau', 'apa', 'yg', 'harus', 'di', 'lakukan', 'guru', 'ke', 'murid']
10.	['saya', 'pikir', 'semua', 'pekerjaan', 'membutuhkan', 'keahlian', 'dan', 'mungkin', 'besaran', 'gaji', 'yg', 'diterima', 'pekerja', 'itu', 'sebanding', 'dengan', 'keterampilan', 'mengatasi', 'seberapa', 'sulit', 'pekerjaan', 'itu', 'di', 'bebankan']

3.2.6 Filtering / Stopwords

Tahap *tokenizing* digunakan dengan membedah anatomi dari suatu kata, setelah itu melanjutkan dengan memeriksa daftar kata yang dinamakan tahap *filtering / stopword*.



Gambar 3.8 Tahap Filtering

Pada tahap ini, dilakukan proses penyaringan kata. Kata-kata yang dipilih dari data adalah kata-kata yang kurang informatif untuk penelitian, sehingga dengan menghilangkannya dapat mempermudah klasifikasi hasil data. *filtering* dalam data juga sangat berguna untuk mengelola dan menganalisis data secara efisien.

Kata yang telah dipilih oleh penulis sebagai kata yang dianggap kurang bermakna dalam teks yaitu 'elo', 'gua', 'gue', 'lo', 'lu', 'si', 'user', 'fafifu', 'sok', 'tau', 'goyang', 'gembul', 'aja', 'yg', 'wkwkwkw', 'gt', 'juga', 'hai', 'dg', 'ttp', 'gw', 'dll', 'dg', 'agak', 'hehe', 'sih', 'su', 'wanjir', 'bgt', 'ato', 'nah', 'yo', 'ni', 'ko', 'niy', 'ya', 'dkk',

'tuh', 'bang', 'nya', 'ada', 'saya', 'lagi', 'lg', 'ga', 'aku'. Dibawah ini merupakan data yang telah dilakukan *filtering/stopword*.

Tabel 3.6 Hasil *Filtering/Stopword*

No.	Data yang telah di <i>Stopwords</i>
1.	['gaji', 'guru', 'indonesia', 'dibanding', 'pns', 'gaji', 'bagus', 'buktinya', 'guru', 'mobil', 'rumah', 'layak', 'bilang', 'sepeda', 'ontel', 'bohong', 'kemaruk', 'disamping', 'gaji', 'remunerasi', 'besarnya', 'gajinya', 'diperhatikan', 'adalah', 'penempatannya', 'sekolah', 'kebanyakan', 'guru', 'didaerah', 'kekurangan', 'guru', 'mestinya', 'ditempatkan', 'diseluruh', 'wilayah', 'indonesia', 'perjanjian', 'dikota']
2.	['salah', 'terpenting', 'guru', 'memiliki', 'gairah', 'pendidikan', 'tulus', 'membantu', 'siswa', 'belajar', 'berkembang']
3.	['tolong', 'bahas', 'angka', 'kemiskinan', 'indonesia', 'tingkat', 'pendidikan']
4.	['profesi', 'guru', 'tapi', 'system', 'negara', 'aneh']
5.	['guru', 'mengajar', 'mendidik', 'orang', 'mengajar', 'guru', 'mendidik']
6.	['evan', 'menyimak', 'terima', 'kasih', 'menyajikan', 'konten', 'rupa', 'sistematis', 'runut', 'referensi', 'referensinya', 'sukses', 'tim', 'persen']
7.	['perubahan', 'sistim', 'pendidikan', 'indonesia', 'merubah', 'perubahan', 'maju']

8.	[‘rill’, ‘recruit’, ‘guru’, ‘honorer’, ‘kampung’, ‘mudah’, ‘klo’, ‘orang’, ‘pas’, ‘sd’, ‘tukang’, ‘jaga’, ‘sekolah’, ‘jualan’, ‘kantin’, ‘guru’, ‘olahraga’, ‘rada’, ‘toxic’, ‘ngomong’, ‘kasar’, ‘menjadikan’, ‘murid’, ‘ngikutin’, ‘ngomong’, ‘kasar’]
9.	[‘karna’, ‘menggurui’, ‘benci’, ‘guru’, ‘bener’, ‘menggurui’, ‘mengerti’, ‘murid’, ‘lakukan’, ‘guru’, ‘murid’]
10.	[‘pikir’, ‘pekerjaan’, ‘membutuhkan’, ‘keahlian’, ‘dan’, ‘besaran’, ‘gaji’, ‘diterima’, ‘pekerja’, ‘sebanding’, ‘keterampilan’, ‘mengatasi’, ‘sulit’, ‘pekerjaan’, ‘bebankan’]

3.2.7 *Stemming*

Setelah dilakukannya *stopword* berlanjut pada proses *stemming*. *Stemming* merupakan tahap dari penghapusan imbuhan sehingga kata akan berubah ke bentuk dasarnya. Berikut ini merupakan langkah dari proses *stemming*.



Gambar 3.9 Tahap *Stemming*

Diawali dari proses input data hasil *stopword* lalu masuk pada proses memeriksa daftar kata *stemming*, proses ini bertujuan untuk mengubah kata berimbuhan menjadi bentuk kata dasarnya jadi menghilangkan kata imbuhan seperti akhiran (-nya, -lah, -kan). Proses *stemming* menggunakan *library* sastrawi untuk mengembalikan sebuah kata ke bentuk dasar. Setelah selesai hasil *stemming* akan menghasilkan data seperti pada tabel dibawah ini.

Tabel 3.7 Hasil *Stemming*

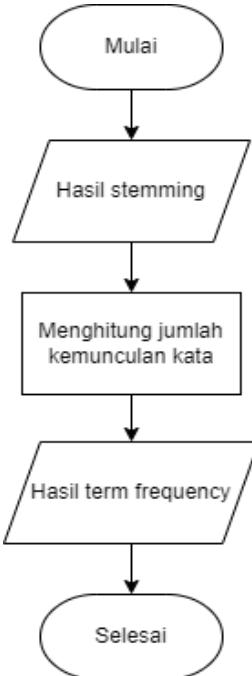
No.	Data hasil <i>Stopwords</i>	Data sesudah di <i>Stemming</i>
1.	['gaji', 'guru', 'indonesia', 'dibanding', 'pns', 'gaji', 'bagus', 'buktinya', 'guru', 'mobil', 'rumah', 'layak', 'bilang', 'sepeda', 'ontel', 'bohong', 'kemaruk', 'disamping', 'remunerasi', 'besarnya', 'gajinya', 'diperhatikan', 'adalah', 'penempatannya', 'sekolah', 'kebanyakan', 'guru', 'didaerah', 'kekurangan', 'guru', 'mestinya', 'ditempatkan', 'diseluruh', 'wilayah', 'indonesia', 'perjanjian', 'dikota']	['gaji', 'guru', 'indonesia', 'banding', 'pns', 'gaji', 'bagus', 'bukti', 'guru', 'mobil', 'rumah', 'layak', 'bilang', 'sepeda', 'ontel', 'bohong', 'kemaruk', 'samping', 'gaji', 'remunerasi', 'besar', 'gaji', 'perhati', 'adalah', 'penempatan', 'sekolah', 'banyak', 'guru', 'daerah', 'kurang', 'guru', 'mesti', 'tempat', 'seluruh', 'wilayah', 'indonesia', 'janji', 'kota']
2.	['salah', 'terpenting', 'guru', 'memiliki', 'gairah', 'pendidikan', 'siswa', 'belajar', 'kembang']	['salah', 'penting', 'guru', 'memiliki', 'gairah', 'pendidikan', 'tulus', 'bantu', 'siswa', 'belajar', 'kembang']

	'tulus', 'membantu', 'siswa', 'belajar', 'berkembang']	
3.	['tolong', 'bahas', 'angka', 'kemiskinan', 'indonesia', 'tingkat', 'pendidikan']	['tolong', 'bahas', 'angka', 'miskin', 'indonesia', 'tingkat', 'didik']
4.	['profesi', 'guru', 'tapi', 'system', 'negara', 'aneh']	['profesi', 'guru', 'system', 'negara', 'aneh']
5.	['guru', 'mengajar', 'mendidik', 'orang', 'mengajar', 'guru', 'mendidik']	['guru', 'ajar', 'didik', 'orang', 'ajar', 'guru', 'didik']
6.	['evan', 'menyimak', 'terima', 'kasih', 'menyajikan', 'konten', 'rupa', 'sistematis', 'runut', 'referensi', 'referensinya', 'sukses', 'tim', 'persen']	['evan', 'simak', 'terima', 'kasih', 'saji', 'konten', 'rupa', 'sistematis', 'runut', 'referensi', 'referensi', 'sukses', 'tim', 'persen']
7.	['perubahan', 'sistim', 'pendidikan', 'indonesia', 'merubah', 'perubahan', 'maju']	['ubah', 'sistim', 'didik', 'indonesia', 'rubah', 'ubah', 'maju']
8.	['rill', 'recruit', 'guru', 'honor', 'kampung', 'mudah', 'klo', 'orang', 'pas', 'sd', 'tukang', 'jaga', 'sekolah', 'jualan', 'kantin', 'guru', 'olahraga',	['rill', 'recruit', 'guru', 'honor', 'kampung', 'mudah', 'klo', 'orang', 'pas', 'sd', 'tukang', 'jaga', 'sekolah', 'jual', 'kantin', 'guru', 'olahraga', 'rada', 'toxic',

	'rada', 'toxic', 'ngomong', 'kasar', 'menjadikan', 'murid', 'ngikutin', 'ngomong', 'kasar']	'ngomong', 'kasar', 'jadi', 'murid', 'ngikutin', 'ngomong', 'kasar']
9.	['karna', 'menggurui', 'benci', 'guru', 'bener', 'bener', 'menggurui', 'mengerti', 'murid', 'lakukan', 'guru', 'murid']	['karna', 'guru', 'benci', 'guru', 'bener', 'guru', 'erti', 'murid', 'laku', 'guru', 'murid']
10.	['pikir', 'pekerjaan', 'membutuhkan', 'keahlian', 'dan', 'besaran', 'gaji', 'diterima', 'pekerja', 'sebanding', 'keterampilan', 'mengatasi', 'sulit', 'pekerjaan', 'bebankan']	['pikir', 'kerja', 'butuh', 'ahli', 'besar', 'gaji', 'terima', 'kerja', 'banding', 'terampil', 'atas', 'sulit', 'kerja', 'beban']

3.2.8 Term Frequency

Term Frequency (TF) adalah banyaknya kemunculan kata dari data komentar youtube satu persen. Jika dari komentar yang dipilih muncul suatu kata terus menerus, maka pengaruh dari kata itu sangatlah penting bagi data dokumen tersebut. Kalau sebaliknya maka berarti pengaruh kata tersebut pada suatu data dokumen ialah kecil atau tidak terlalu berpengaruh. Berikut ini ialah langkah dalam melakukan *Term Frequency*.



Gambar 3.11 Tahap *Term Frequency*

Proses ini dimulai dari memasukkan hasil *stemming* lalu memproses perhitungan jumlah kemunculan kata pada kolom text yang akan terbagi 3 TF yaitu, hasil TF Positif, TF Negatif, dan TF Netral.

```

# Menghitung Term Frequency (TF) untuk teks dengan sentimen tertentu
vectorizer = CountVectorizer()
X = vectorizer.fit_transform(filtered_texts)

# Mendapatkan daftar term frequency untuk setiap kata
term_frequency_df = pd.DataFrame(X.toarray(), columns=vectorizer.get_feature_names_out())

# Menghitung total term frequency untuk setiap kata
total_term_frequency = term_frequency_df.sum(axis=0)

# Menampilkan hasil term frequency tiap kata untuk teks dengan sentimen tertentu
print(f"Term Frequency (TF) untuk Setiap Kata dalam Teks {sentiment.capitalize()}:")
print(total_term_frequency)
print() # Untuk spasi antar keluaran
  
```

Gambar 3.10 Code Python *Term Frequency*

Setelah mendapatkan daftar *term frequency* untuk setiap kata pada text sentimen, hasil *term frequency* dari setiap kata akan menunjukkan berapa kali kata-kata tersebut muncul dalam teks tersebut. *Term frequency* membantu untuk memahami kata-kata mana yang paling sering digunakan pada teks sentimen. Dibawah ini merupakan contoh kemunculan *term frequency*.

Tabel 3.8 Hasil *Term Frequency*

No.	Kata	Positif	Negatif	Netral
1.	abdi	2	-	-
2.	ajar	15	14	82
3.	didik	25	18	85
4.	fakta	1	1	8
5.	batas	1	-	4
6.	guru	141	65	555
7.	cabang	-	1	-
8.	calon	1	1	7
9.	cocok	2	-	2
10.	pns	2	-	-

Pada contoh diatas terdapat kata “pns” yang didapat dari komentar “guru kalo pns swasta kyk cocok guru honorer” terdapat kemunculan kata “pns” pada komentar positif dengan kemunculan sekali lalu terdapat kata “guru” yang kemunculannya dua kali, kata “swasta” muncul sekali, kata “cocok” muncul sekali, kata “kyk” muncul sekali, dan kata “honorer” muncul sekali.

3.2.9 TF-IDF

Term Frequency–Invert Document Frequency (TF-IDF) merupakan tahap pengukuran seberapa penting suatu kata pada dokumen. TF-IDF dilakukan menggunakan *TfidfVectorizer* yaitu library *scikit-learn* (*sklearn*). Penggunaan *TfidfVectorizer* proses analisa. *TfidfVectorizer* bekerja dengan cara mengubah teks

menjadi representasi numerik yang berdasarkan pada frekuensi kata di dalam dokumen (TF) dan frekuensi kata di seluruh korpus (IDF).

```
# Inisialisasi TF-IDF Vectorizer dengan jumlah fitur yang sama untuk semua kategori
tfidf_vectorizer = TfidfVectorizer(max_features=1000) # Sesuaikan max_features sesuai kebutuhan, di

# Gabungkan semua teks menjadi satu untuk melakukan fit_transform sekali saja
all_texts = positif_texts + negatif_texts + netral_texts
tfidf_all = tfidf_vectorizer.fit_transform(all_texts)

# Pisahkan kembali berdasarkan kategori sentimen
positif_tfidf = tfidf_all[:len(positif_texts)]
negatif_tfidf = tfidf_all[len(positif_texts):len(positif_texts) + len(negatif_texts)]
netral_tfidf = tfidf_all[len(positif_texts) + len(negatif_texts):]

# Buat DataFrame untuk masing-masing kategori
positif_df = pd.DataFrame(positif_tfidf.toarray(), columns=tfidf_vectorizer.get_feature_names_out())
negatif_df = pd.DataFrame(negatif_tfidf.toarray(), columns=tfidf_vectorizer.get_feature_names_out())
netral_df = pd.DataFrame(netral_tfidf.toarray(), columns=tfidf_vectorizer.get_feature_names_out())
```

Gambar 3.12 Code Python TF-IDF

TF-IDF *vectorization* pada data teks dari file excel dan menyimpan hasilnya ke dalam file excel bernama hasil_tfidf. *Fit_transform* digunakan untuk melatih model TF-IDF dengan mengubah suatu teks menjadi *metrics vector*.

Tabel 3.9 Hasil TF-IDF

Dokumen	TF-IDF				
	ajar	didik	fakta	batas	guru
Dokumen 1	0,084026	0,239211	0	0,155732	0,162014
Dokumen 2	0	0,193742	0	0	0,098414
Dokumen 3	0	0,157864	0	0	0,120284
Dokumen 4	0	0,237773	0	0	0,06039
....
Dokumen 103	0	0	0	0	0,068748

Dokumen 104	0	0	0	0	0,102652
Dokumen 105	0	0	0	0	0,049891
Dokumen 106	0,678476	0	0	0	0
Dokumen 107	0,162097	0,102548	0	0	0,052091
....
Dokumen 509	0	0	0,469988	0	0

3.3 Perancangan Model Klasifikasi

Data set akan dibagi sesuai dengan tujuan yang diinginkan. Pada tahap ini Data set akan dibagi menjadi 2 bagian, yaitu data latih dan data uji. Pembagian ke dalam data latih diperlukan agar model dapat mempelajari pola dan mengelompokkan data baru dengan benar. Lalu data uji berfungsi untuk mengetahui akurasi dari model yang dibuat. Setelah data dilatih selesai maka dilakukan pengujian menggunakan data uji dengan Metode klasifikasinya *Naïve Bayes Classifier*. Setelah proses pelatihan selesai, model akan diuji menggunakan data uji untuk mengukur seberapa baik model dapat mengenali dan mengklasifikasikan data baru. Metode klasifikasinya *Naïve Bayes Classifier* sebuah algoritma yang sederhana namun efektif untuk klasifikasi, terutama pada tugas-tugas analisis teks.

3.4 Evaluasi Model Klasifikasi

Pengevaluasian kinerja *Naïve Bayes Classifier* dilakukan dengan penggunaan *confusion matriks*, caranya membandingkan hasil klasifikasi dengan data klasifikasi yang sebenarnya. Tujuan dari adanya evaluasi model adalah menghasilkan *confusion matriks* yang berukuran 3x3, dan mengetahui seberapa akurat model dalam mengklasifikasikan sentimen teks.