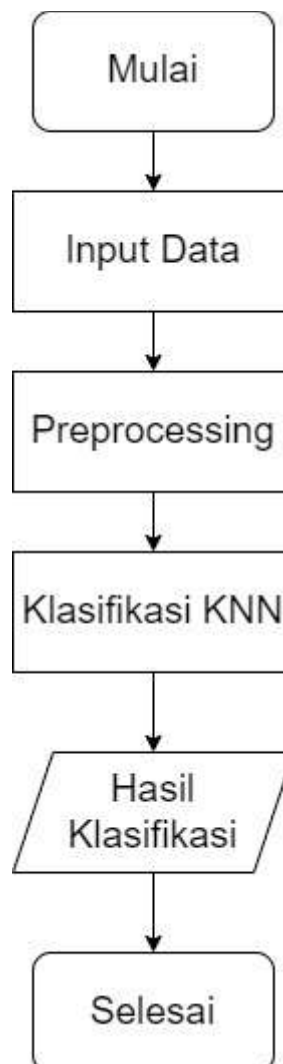


BAB III METODOLOGI PENELITIAN

3.1 Alur Penelitian

Setelah melakukan pengumpulan data dan menentukan metode selanjutnya adalah tinggal melakukan uji coba dari hasil data yang sudah dikumpulkan. Selanjutnya data dari hasil penelitian akan dianalisis dengan cara melakukan verifikasi dan validasi data hingga akan menghasilkan output analisis yang memiliki tingkat akurasi yang tinggi.



Gambar 3.1 Alur Penelitian Analisis Sentimen

3.2 Implementasi Penelitian

Selama meneliti peneliti memerlukan beberapa perangkat lunak (*software*) dan perangkat keras (*hardware*) yang diperlukan untuk mencapai dan membantu penelitian untuk mencapai tujuan yang sudah ditetapkan, diantaranya adalah:

- OS Windows 10
- Aplikasi Jupyter Notebook
- Laptop Lenovo AMD Athlon Gold 8 GB
- Bahasa Pemrograman Python

3.2 Pengumpulan Data

Selama pengumpulan data, peneliti akan mengumpulkan data dengan mencari referensi jurnal yang terkait dengan topik penelitian ini. Pengumpulan data juga dilakukan dengan cara menangkap tweet dengan cara mendapatkan data secara langsung (Crawl) Twitter menggunakan API (Application Interface) pada Twitter.(Legianto, 2019)



Gambar 3.2.1 Pengumpulan Data menggunakan Rapidminer

3.3 Preprocessing

Setelah mendapatkan data, yang harus Anda lakukan adalah membersihkan data. Pembersihan dilakukan dengan menghilangkan noise, menghapus data duplikat, memeriksa data untuk ketidakkonsistenan, dan memperbaiki kesalahan dalam data, seperti kesalahan ketik.

Pada tahap ini data akan dibagi menjadi data testing dan data training dengan rasio 80% untuk data testing dan 20% data training dan didapatkan hasil :

Tabel 3.3.1 Pembagian Data Testing dan Data Training

Jumlah Data (N)	Data Testing (N x 20%)	Data Training (N x 80%)
704	141	563

3.3.1 Cleansing Data

Tahap Cleansing data dilakukan sebagai tahap awal yang sangat penting dalam penelitian ini. Hal ini dikarenakan data yang didapatkan dari Twitter masih dalam bentuk teks yang tidak sesuai kaidah dan tidak lengkap. Normalisasi adalah proses penskalaan nilai atribut dari data sehingga terletak pada rentang tertentu (Nasution et al., 2019).

Selain itu pada tahap normalisasi juga dilakukan secara *Cleaning data* menggunakan pemrograman Python, hal ini bertujuan untuk memastikan isi dari data tersebut. *Cleaning data* yang akan menghapus atribut-atribut yang tidak diperlukan di dalam data seperti, tanda baca, emot ikon, angka dan lain-lain. Atribut-atribut yang akan dihapus contohnya “@[A-Za-z0-9]+, [0-9]+, #, [^\w], ‘_’, [\n]+, :, ‘RT[\s]+, ^https?:\V\.[\r\n], ^http?:\V\.[\r\n]”.

```

def cleanTweet(text):
    text = re.sub(r'@[A-Za-z0-9]+', '', text)
    text = re.sub(r'[0-9]+', '', text)
    text = re.sub(r'#', '', text)
    text = re.sub(r'^\w', ' ', text)
    text = re.sub(r'_', ' ', text)
    text = re.sub(r'[\n]+', '', text)
    text = re.sub(r':', '', text)
    text = re.sub(r'RT[\s]+', '', text)

    return text
df['clean_tweets'] = df['Text'].apply(cleanTweet)

df.head()

```

Gambar 3.3.1.1 Cleansing Data dengan Python

Tabel 3.3.1.1 Hasil Sebelum dan sesudah di Cleansing

No	Data Awal	Cleansing data
1.	Pemerintah melobi berbagai pihak agar RKUHP segera disahkan. Pasal yang mengancam demokrasi dan hak privat dipertahankan. #MajalahTempo <a 1552487543077163008"="" href="https://t.co/fZyE7WMTEo,">https://t.co/fZyE7WMTEo,"1552487543077163008	Pemerintah melobi berbagai pihak agar RKUHP segera disahkan Pasal yang mengancam demokrasi dan hak privat dipertahankan https://t.co/fZyE7WMTEo "1552487543077163008"
2.	Pada 6 Juli lalu, Pemerintah melalui Kementerian Hukum dan HAM resmi menyerahkan draf RKUHP ke DPR. Penyusunan draf RKHUP yang sudah berlangsung sejak lama dan sempat batal disahkan pada 2019 lalu ini, menimbulkan kritik dari banyak pihak. <a 1549344544243986432"="" href="https://t.co/NEEjKm10d4,">https://t.co/NEEjKm10d4,"1549344544243986432"	Pada Juli lalu Pemerintah melalui Kementerian Hukum dan HAM resmi menyerahkan draf RKUHP ke DPR Penyusunan draf RKHUP yang sudah berlangsung sejak lama dan sempat batal disahkan pada

		2019 lalu ini menimbulkan kritik dari banyak pihak https://t.co/NEEjKm10d4 “1549344544243986432”
--	--	--

3.3.2 Labeling Data

Labeling data adalah tahap memberikan label pada setiap data yang sudah dibersihkan atau telah melalui Cleansing data. Dalam penelitian ini data akan dilabeli melalui Natural Language Processing yang ada di Library Python yaitu Textblob.

Proses Labeling melalui Textblob hanya bisa dilakukan dalam Bahasa Inggris jadi dalam melakukan Labeling data yang sudah melalui tahap cleansing data akan diterjemahkan terlebih dahulu kedalam Bahasa Inggris sebelum melakukan Labeling pada data. Proses Labeling juga dilakukan dengan berkonsultasi kepada ahli Bahasa yaitu Bapak Dr. Muhammad Sukirlan, M.A. selaku kepala UPT Bahasa di Unila, dan didapatkan hasil bahwa:

Tabel 3.3.2.1 Penjelasan Masing-Masing Label

No.	Sentimen	Penjelasan
1.	Negatif	Pernyataan yang sangat tidak mendukung terhadap konten yang dimaksudkan.
2.	Netral	Pernyataan yang tidak termasuk dalam mendukung dan setuju terhadap konten serta tidak termasuk kedalam pernyataan yang menentang atau tidak mendukung konten.
3.	Positif	Pernyataan yang sangat mendukung atau mengajak terhadap konten yang dimaksudkan.

Menurut Bapak Dr. Muhammad Sukirlan, M.A. labeling dilakukan dengan

melihat kata kerja, kata bantu, kata sifat dan juga kata sambung di kalimatnya, contoh beberapa kata yang bisa diindikasikan sebagai kata yang negatif atau positif:

Tabel 3.3.2.2 Contoh Kata Negatif dan Kata Negatif

Positif	Negatif
Senang	Benci
Ayo	Tua
Dukung	Tidak
Ajak	Tertindas
Percaya	Khawatir
Perlu	Bukan
Menaati	Melanggar

Beliau juga mengatakan bahwa jika ada kata Positif namun diawali dengan kata “Tidak” maka kata tersebut akan berlabel Negatif, begitu juga sebaliknya jika ada kata Negatif yang diawali dengan kata “Tidak” maka kata tersebut akan berlabel Positif, sedangkan untuk kata Netral beliau mengatakan bahwa kalimatnya akan mengandung keduanya atau tidak mengandung keduanya dan juga bisa berupa pertanyaan, berikut contoh kalimatnya:

Tabel 3.3.2.2 Cara Melabeli Kalimat

Kalimat	Sentimen
Senang sekali membaca tulisan dari tentang RKHUP dan perjalanannya dari rumah ke rumah seperti lagunya	Positif
Harkristuti Harkrisnowo berpendapat bahwa KUHP yang diterapkan sekarang sudah terlalu tua dan bukan merupakan buatan Indonesia	Negatif
Mana suaranya yang sepakat mempertemukan dan untuk debat RKUHP	Netral

3.3.3 Pembobotan Kata

Pada tahap pembobotan kata atau *Terms Weight* ini akan dilakukan perhitungan secara otomatis menggunakan rapidminer. Namun pada kali ini akan memberikan contoh dan Contoh akan dibagi menjadi 4 dimana 3 adalah data training dan 1 adalah data testing. Berikut adalah perhitungan manual dari pembobotan kata, yaitu:

Tabel 3.3.3.1 Contoh Data Pembobotan Kata

Ket	Tweet	Sentimen
D1	gagalkan rkuhp uu kolonial gaya oligarki hentikan periode oligarki	Negatif
D2	hak publik untuk bersuara dan ikut terlibat dalam proses perumusan rkuhp ini	Positif
D3	senang sekali membaca tulisan dari tentang rkhup dan perjalanannya dari rumah ke rumah seperti lagunya	Positif
D4	rkuhp memudahkan masyarakat mencari pekerjaan tapi banyak pasal karet	?

Selanjutnya melakukan tokenisasi dan lakukan *stopwords*, tokenisasi dilakukan untuk memisahkan kalimat menjadi kata-kata dan *stopwords* digunakan untuk membuang kata-kata yang dirasa memiliki makna yang kurang berarti seperti “yang”, “dan”, “atau” dan lain-lain. Ini merupakan hasilnya:

Tabel 3.3.3.2 Data Pembobotan Kata Tf-Idf

Term	Tf					Idf	Wt = Tf.Idf			
	D1	D2	D3	D4	df	log(n/df)	D1	D2	D3	D4
gagalkan	1	0	0	0	1	0,60205999	0,60205999	0	0	0
rkuhp	1	1	1	1	4	0	0	0	0	0
uu	1	0	0	0	1	0,60205999	0,60205999	0	0	0
kolonial	1	0	0	0	1	0,60205999	0,60205999	0	0	0
gaya	1	0	0	0	1	0,60205999	0,60205999	0	0	0
oligarki	2	0	0	0	2	0,30103	0,60205999	0	0	0
hentikan	1	0	0	0	1	0,60205999	0,60205999	0	0	0
periode	1	0	0	0	1	0,60205999	0,60205999	0	0	0
hak	0	1	0	0	1	0,60205999	0	0,60205999	0	0
publik	0	1	0	0	1	0,60205999	0	0,60205999	0	0
untuk	0	1	0	0	1	0,60205999	0	0,60205999	0	0
bersuara	0	1	0	0	1	0,60205999	0	0,60205999	0	0
ikut	0	1	0	0	1	0,60205999	0	0,60205999	0	0
terlibat	0	1	0	0	1	0,60205999	0	0,60205999	0	0
proses	0	1	0	0	1	0,60205999	0	0,60205999	0	0
perumusan	0	1	0	0	1	0,60205999	0	0,60205999	0	0
senang	0	0	1	0	1	0,60205999	0	0	0,60205999	0
membaca	0	0	1	0	1	0,60205999	0	0	0,60205999	0
tulisan	0	0	1	0	1	0,60205999	0	0	0,60205999	0
tentang	0	0	1	0	1	0,60205999	0	0	0,60205999	0
perjalanannya	0	0	1	0	1	0,60205999	0	0	0,60205999	0
rumah	0	0	1	0	1	0,60205999	0	0	0,60205999	0
lagunya	0	0	1	0	1	0,60205999	0	0	0,60205999	0
memudahkan	0	0	0	1	1	0,60205999	0	0	0	0,60205999
masyarakat	0	0	0	1	1	0,60205999	0	0	0	0,60205999
mencari	0	0	0	1	1	0,60205999	0	0	0	0,60205999
pekerjaan	0	0	0	1	1	0,60205999	0	0	0	0,60205999
pasal	0	0	0	1	1	0,60205999	0	0	0	0,60205999
karet	0	0	0	1	1	0,60205999	0	0	0	0,60205999

Kata “gagalkan” memiliki *df* (*document frequency*) sebanyak 1 dari 4 dokumen yang ada. Jadi dalam menentukan *Idf* (*inverse document frequency*) adalah:

$$\text{Idf} = \log (4/1)$$

$$\text{Idf} = \log (4)$$

$$\text{Idf} = 0,60205999$$

Setelah mendapatkan nilai *Idf*, dapat dilanjutkan ketahap selanjutnya dengan mengkalikan *Idf* dan *Tf* (*Terms frequency*), dengan nilai *Tf* “gagalkan” di masing-masing dokumen $D1 = 1$, $D2 = 0$, $D3=0$, $D4=0$, maka dapat dihasilkan *Wt* (*Weight terms*) sebagai berikut :

$$\text{Wt} = \text{Tf} . \text{Idf}$$

$$D1 =1$$

$$\text{Wt} = 1. 0,60205999$$

$$\text{Wt} = 0,60205999$$

$$D1 =2$$

$$\text{Wt} = 0. 0,60205999$$

$$\text{Wt} = 0$$

$$D1 =3$$

$$\text{Wt} = 0. 0,60205999$$

$$\text{Wt} = 0$$

$$D1 =5$$

$$\text{Wt} = 0. 0,60205999$$

$$\text{Wt} = 0$$

Data hasil dari pembobotan kata akan digunakan untuk mencari *Similiarity* atau persamaan menggunakan metode K-NN

0	0	0
0	0	0
0	0	0
0	0	0
0	0	0
0	0	0

Perkalian Sklar yang dimaksudkan untuk mencari nilai $\sum_k(d_{ik} \cdot d_{kj})$ pada rumus similiarity, dimana d_{ik} adalah bobot data dokumen *Testing* (D1, D2, D3) d_{kj} adalah bobot data dokumen *Training* (D4). Perkalian ini pasangan untuk masing-masing nilai *Weight Terms* atau bobot kata, contoh perhitungan manual nya :

D1: “gagalkan” = 0,60205999

D2: “gagalkan” = 0

D3: “gagalkan” = 0

D4: “gagalkan” = 0

$$\sum_k (d_{ik} \cdot d_{kj})$$

D1 = bobot D1 . bobot D5

D1 = 0,60205999 . 0

D1 = 0

D2 = bobot D3 . bobot D5

D2 = 0 . 0

D2 = 0

D3 = bobot D3 . bobot D5

D3 = 0 . 0

D3 = 0

Selanjutnya adalah menghitung vector Panjang setiap dokumen dengan rumus :

$$\sqrt{\sum_k d^2 ik} \sqrt{\sum_k d^2 jk}$$

Tabel 3.4.1 Perkalian Panjang Vektor

Panjang Vektor			
D1	D2	D3	D4
0,36247623	0	0	0
0	0	0	0
0,36247623	0	0	0
0,36247623	0	0	0
0,36247623	0	0	0
0,36247623	0	0	0
0,36247623	0	0	0
0,36247623	0	0	0
0	0,36247623	0	0
0	0,36247623	0	0
0	0,36247623	0	0
0	0,36247623	0	0
0	0,36247623	0	0
0	0,36247623	0	0
0	0,36247623	0	0
0	0,36247623	0	0
0	0,36247623	0	0
0	0	0,36247623	0
0	0	0,36247623	0
0	0	0,36247623	0
0	0	0,36247623	0
0	0	0,36247623	0
0	0	0,36247623	0
0	0	0,36247623	0
0	0	0	0,36247623
0	0	0	0,36247623
0	0	0	0,36247623
0	0	0	0,36247623
0	0	0	0,36247623
0	0	0	0,36247623
2,53733362	2,89980985	2,53733362	2,17485738

1,59290100	1,70288280	1,59290100	1,47473976
------------	------------	------------	------------

Panjang vektor :

$$D1 : \sqrt{0,60205999^2 \times 7}$$

$$D1 : \sqrt{2,53733362}$$

$$D1 : 1,59290100$$

$$D2 : \sqrt{0,60205999^2 \times 8}$$

$$D2 : \sqrt{2,89980985}$$

$$D2 : 1,70288280$$

$$D3 : \sqrt{0,60205999^2 \times 7}$$

$$D3 : \sqrt{2,53733362}$$

$$D3 : 1,59290100$$

$$D4 : \sqrt{0,60205999^2 \times 6}$$

$$D4 : \sqrt{2,17485738}$$

$$D4 : 1,47473976$$

Jika semua sudah didapatkan maka data tersebut dapat dimasukkan kedalam rumus yang utuh :

$$\text{Cos}(D4,D1) = 0 / (1,47473976*1,59290100) = 0$$

$$\text{Cos}(D4,D2) = 0 / (1,47473976*1,70288280) = 0$$

$$\text{Cos}(D4,D3) = 0 / (1,47473976*1,59290100) = 0$$

Nilai kedekatan vector yang dimiliki oleh D1, D2, D3 terhadap D4 memiliki nilai yang sama yaitu 0, maka dapat disimpulkan bahwa D4 bukan termasuk dalam sentiment “Negatif” maupun “Positif” atau D4 termasuk dalam sentiment “Negatif” maupun “Positif” yang mana itu berarti D4 adalah “Netral”.