

BAB IV HASIL DAN PEMBAHASAN

4.1 Hasil Pengujian Perangkat Lunak Labeling

Penelitian ini akan menggunakan perangkat lunak berbasis python dalam melabali data. Berikut adalah langkah-langkahnya:

- a. *Install library translate* yang ada di *python*, hal ini dikarenakan *library Textblob* yang nanti akan digunakan hanya bisa menggunakan bahasa inggris dalam pengklasifikasian.

```
In [1]: from googletrans import Translator

In [2]: translator = Translator()
translations = translator.translate(['The quick brown fox', 'jumps over', 'the lazy dog'], dest='ko')
for translation in translations:
    print(translation.origin, ' -> ', translation.text)

The quick brown fox -> 빠른 갈색 여우
jumps over -> 뛰어 넘다
the lazy dog -> 게으른 개
```

===== BATAS =====

Gambar 4.1.1 Proses *install library translate*.

- b. Import library yang akan digunanakan nanti, seperti :
 - Pandas
 - Numpy
 - Seaborn
 - String
 - Nltk
 - Re(regex)
 - Textblob
 - Collections dll.

- c. Selanjutnya import data yang sudah didapatkan dari *crawling data twitter*.

```
In [3]: import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt
from sklearn.feature_extraction.text import CountVectorizer
import nltk
import string
import re
from textblob import TextBlob
from collections import Counter

def load_data():
    data = pd.read_csv('Data Jadi/Data Awal A.csv')
    return data

tweet_df = load_data()
tweet_df.head()

df = pd.DataFrame(tweet_df[['Id', 'Text']])
```

Gambar 4.1.2 Proses *import data twitter*.

- d. Pada tahap selanjutnya, data yang sudah di impor akan di bersihkan atau masuk ke tahap cleansing data, dengan cara menghapuskan simbol-simbol dan juga angka.

```
In [6]: 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()
```

Out[6]:

	Id	Text	clean_tweets
0	1556183140679966208	Senang sekali membaca tulisan @sksrwy dan ..	Senang sekali membaca tulisan dari henta ..
1	1554822444296196097	Kamu perlu mengantongi pasai-pasai krusial dal ..	Kamu perlu mengantongi pasai pasai krusial dal ..
2	1556538323673677824	Mana suaranya yang sepakat memperlemukan @eddy ..	Mana suaranya yang sepakat memperlemukan dan ..
3	1556624757034188890	KAKANWIL KEMENKUMHAM JATENG A YUSPAHRUDDIN: Ka ..	KAKANWIL KEMENKUMHAM JATENG A YUSPAHRUDDIN Ka ..
4	1556624708045822720	KAKANWIL KEMENKUMHAM JATENG A YUSPAHRUDDIN: Ka ..	KAKANWIL KEMENKUMHAM JATENG A YUSPAHRUDDIN Ka ..

Gambar 4.1.3 Proses *cleansing data*

- e. Data yang sudah bersih pada filed `clean_tweets` akan diubah menjadi bahasa inggris menggunakan library `translate` yang ada di python agar nanti bisa dibaca oleh library `textblob`.

```
In [7]: translator = Translator()
from_text = df['clean_tweets']
to_text = []
for text in from_text:
    translations = translator.translate(text, dest='en')
    to_text.append(translations.text)
print(to_text)

["It's really nice to read articles about RKUHP and the journey from house to house like the song https t co azqP8iIR", "You need to pocket the crucial articles in this RKUHP because soon the government will hold an open discussion There are problems tic articles and we must not be careless Mata Najwa https t co HuarBiliv", "Where is the voice that agrees to bring together and debate the RKUHP?", "KKANWIL KKANWIL KEMENKUMHAM CENTRAL JAVA A YUSPAHRUDDIN kakawil Kemerkumham Central Java Invites His Staff to Socialize RKUHP https t co eAsQWRAI", "KKANWIL KKANWIL KEMENKUMHAM CENTRAL JAVA A YUSPAHRUDDIN kakawil Kemerkumham Central Java Invites His Staff to Socialize RKUHP https t co rDatpSLn", "KEMANWIL KEMENKUMHAM CENTRAL JAVA A YUSPAHRUDDIN Head of Regional Office of Kemerkumham Central Java Invites His Staff to Socialize RKUHP https t co LtASLQmK", "KKANWIL KKANWIL KEMENKUMHAM CENTRAL JAVA A YUSPAHRUDDIN kakawil Kemerkumham Central Java Invites His Staff to Socialize RKUHP https t co qVqcrakI", "KKANWIL KKANWIL KEMENKUMHAM CENTRAL JAVA A YUSPAHRUDDIN kakawil Kemerkumham Central Java Invites His Staff to Socialize RKUHP https t co H5BlaQVnK", "KKANWIL KKANWIL KEMENKUMHAM CENTRAL JAVA A YUSPAHRUDDIN kakawil Kemerkumham Central Java Invites His Staff to Socialize RKUHP https t co BfUwREkE", "KKANWIL KKANWIL KEMENKUMHAM CENTRAL JAVA A YUSPAHRUDDIN kakawil Kemerkumham Central Java Invites His Staff to Socialize RKUHP https t co gIXEN", "KKANWIL KKANWIL KEMENKUMHAM CENTRAL JAVA A YUSPAHRUDDIN Head of Regional Kemerkumham Central Java Invites His Staff to Socialize RKUHP https t co SHWVjluC", "KKANWIL KEMENKUMHAM CENTRAL JAVA A YUSPAHRUDDIN kakawil Kemerkumham Central Java Invites His Staff to Socialize RKUHP https t co IDYkInIAH", "KKANWIL KKANWIL KEMENKUMHAM CENTRAL JAVA A YUSPAHRUDDIN kakawil Kemerkumham Central Java Invites His Staff to Socialize RKUHP https t co qV483rThs", "KKANWIL KKANWIL KEMENKUMHAM CENTRAL JAVA A YUSPAHRUDDIN kakawil Kemerkumham Central Java Invites His Staff to Socialize RKUHP https t co OYVLEZJaI", "KKANWIL KEMENKUMHAM CENTRAL JAVA A YUSPAHRUDDIN Head of Regional Kemerkumham Central Java Invites His Staff to Socialize RKUHP https t co P4u8uQ2h", "IS KKANWIL KEMENKUMHAM CENTRAL JAVA A YUSPAHRUDDIN Gives Public Lecture The Deputy Minister of Law and Human Rights of the Republic of Indonesia Explains the U
```

Gambar 4.1.4 data twitter bersih yang sudah ditranslate

```
In [8]: df['in_english'] = to_text
print(df['in_english'])

0      It's really nice to read articles about RKUHP ...
1      You need to pocket the crucial articles in thi...
2      Where is the voice that agrees to bring togeth...
3      KKANWIL KKANWIL KEMENKUMHAM CENTRAL JAVA A YUS...
4      KKANWIL KKANWIL KEMENKUMHAM CENTRAL JAVA A YUS...
...
1773   Students Hold Action Against RKUHP https t co ...
1774   KUHP Article Insulting the State The Doctrine ...
1775   The government is okay with cursing and compla...
1776   People want to be brought back to the colonial...
1777   KUHP Article Insulting the State The Doctrine ...
Name: in_english, Length: 1978, dtype: object
```

Gambar 4.1.5 proses membuat dataframe pada data twitter

- f. Setelah mendapatkan data yang sudah di ubah ke dalam bahasa inggris, selanjutnya ubah data ke dalam bentuk data frame atau field.

```
In [9]: def cleanTweet(text):
text = re.sub(r'@[A-Za-z0-9]+', '', text)
text = re.sub(r'#[a-z0-9]+', '', text)
text = re.sub(r'http://', '', text)
text = re.sub(r'rt[ ]+', '', text)
text = re.sub(r'[\n]', '', text)
text = re.sub(r'RT[ ]+', '', text)
re.sub(r'https?://.*[\r\n]', '', text, flags=re.MULTILINE)
re.sub(r'^https?://.*[\r\n]', '', text, flags=re.MULTILINE)

return text
# Lets apply to our data frame
df['clean_tweets_english'] = df['in_english'].apply(cleanTweet)
# Lets take a look
df.head()
```

Out[9]:	id	Text	clean_tweets	in_english	clean_tweets_english
0	155018314087090208	Senang sekali membaca tulisan @winny dan	Senang sekali membaca tulisan dan teks	It's really nice to read articles about RKUHP	It's really nice to read articles about RKUHP
1	1514523444290190091	Kamu perlu mengartikan pasal pasal itu di	Kamu perlu mengartikan pasal pasal itu di	You need to pocket the crucial articles in th	You need to pocket the crucial articles in th
2	155658323273677204	Mata sukanya yang enak di matanya dan	Mata sukanya yang enak di matanya dan	Where is the voice that agrees to bring togeth	Where is the voice that agrees to bring togeth
3	1558824757034190060	KKANWIL KEMENKUMHAM JATENG A YUSPAHRUDDIN Ka	KKANWIL KEMENKUMHAM JATENG A YUSPAHRUDDIN Ka	KKANWIL KKANWIL KEMENKUMHAM CENTRAL JAVA A YUS	KKANWIL KKANWIL KEMENKUMHAM CENTRAL JAVA A YUS
4	155882470949822738	KKANWIL KEMENKUMHAM JATENG A YUSPAHRUDDIN Ka	KKANWIL KEMENKUMHAM JATENG A YUSPAHRUDDIN Ka	KKANWIL KKANWIL KEMENKUMHAM CENTRAL JAVA A YUS	KKANWIL KKANWIL KEMENKUMHAM CENTRAL JAVA A YUS

Gambar 4.1.6 proses translate data dari bahasa inggris ke bahasa indonesia

- g. Setelah data siap untuk diolah, maka selanjutnya data yang sudah ada di `clean_tweets_english` akan dilakukan percabangan untuk mengetahui label dari masing-masing file atau isi datanya. Data akan dibagi berdasarkan polarity dan subjectivity yang akan di cabangkan menjadi negatif, netral dan positif yang sudah memiliki ketentuannya seperti “`data < 0 = negatif`”, “`data == 0 =netral`”, “`data > 0 =positif`”.

```
In [10]: analysis = TextBlob('in_english')

def getSubjectivity(text):
    return TextBlob(text).sentiment.subjectivity
def getPolarity(text):
    return TextBlob(text).sentiment.polarity

def analyse(score):
    if score < 0:
        return 'Negative'
    elif score == 0:
        return 'Neutral'
    else:
        return 'Positive'

final_data = pd.DataFrame(df[['clean_tweets_english']])

final_data['clean_tweets'] = df['clean_tweets']
final_data['Subjectivity'] = final_data['clean_tweets_english'].apply(getSubjectivity)
final_data['Polarity'] = final_data['clean_tweets_english'].apply(getPolarity)
final_data['Sentiment'] = final_data['Polarity'].apply(analyse)

final_data.head()
```

```
Out[10]:
```

	clean_tweets_english	clean_tweets	Subjectivity	Polarity	Sentiment
0	It's really nice to read articles about RKNUP	Senang sekali membaca tulisan dari lanta	1.00	0.600000	Positive
1	You need to pocket the crucial articles in thi	Kamu perlu mengartongi pasal pasal kristal dal	0.30	-0.105567	Negative
2	What's the voice that agrees to bring togeth	Mana suaranya yang sepakat memperlemukan dan	0.00	0.000000	Neutral
3	KKANWIL KKANWIL KEMENKUMHAM CENTRAL JAWA A YUS	KAKANWIL KEMENKUMHAM JATENG A YUSPAHRUDDIN Ka	0.25	0.000000	Neutral
4	KKANWIL KKANWIL KEMENKUMHAM CENTRAL JAWA A YUS	KAKANWIL KEMENKUMHAM JATENG A YUSPAHRUDDIN Ka	0.25	0.000000	Neutral

Gambar 4.1.7 proses klasifikasi menggunakan textblob

- h. Setelah data sudah dilabeli maka data dapat di ekspor untuk nanti akan diolah. Data juga dapat dilihat berapa jumlahnya.

```
In [11]: positive_counts = Counter()
negative_counts = Counter()
total_counts = Counter()

for i in range(len(final_data["Sentiment"])):
    for word in final_data["Sentiment"][i].lower().split(" "):
        positive_counts[word]+=1
        total_counts[word]+=1

for i in range(len(final_data["Sentiment"])):
    for word in final_data["Sentiment"][i].lower().split(" "):
        negative_counts[word]+=1
        total_counts[word]+=1
```

```
In [12]: negative_counts.most_common()[0:10]
```

```
Out[12]: [('neutral', 1174), ('negative', 426), ('positive', 378)]
```

```
In [13]: len(final_data["Sentiment"])
```

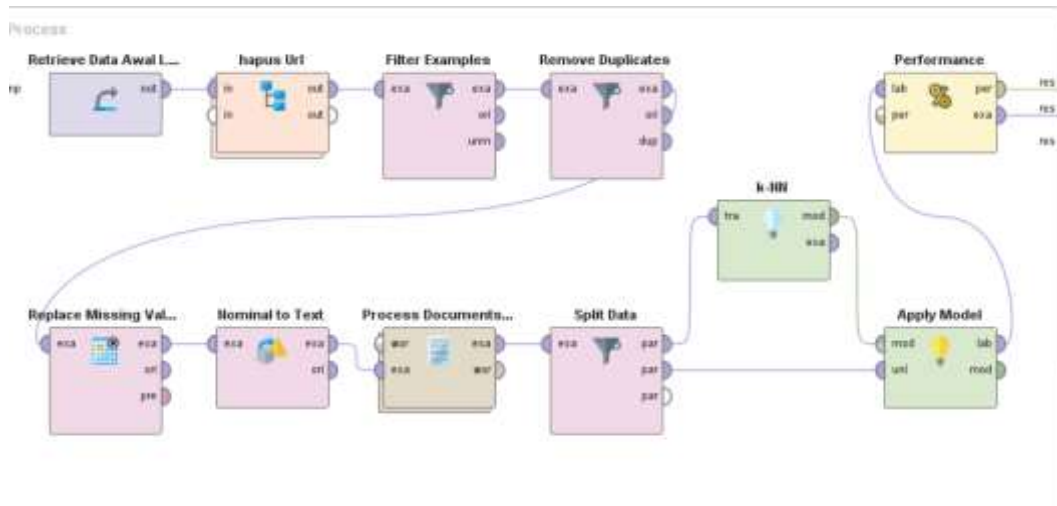
```
Out[13]: 1978
```

```
In [24]: final_data.to_csv("data.csv", index=False)
```

Gambar 4.1.8 proses *eksporting* data

4.2 Implementasi Algoritma K-Nearest Neighbor menggunakan Klasifikasi

Berikut merupakan implementasi Algoritma Klasifikasi K-Nearest Neighbor menggunakan Rapidminer dengan tool *K-Nearest Neighbor* :

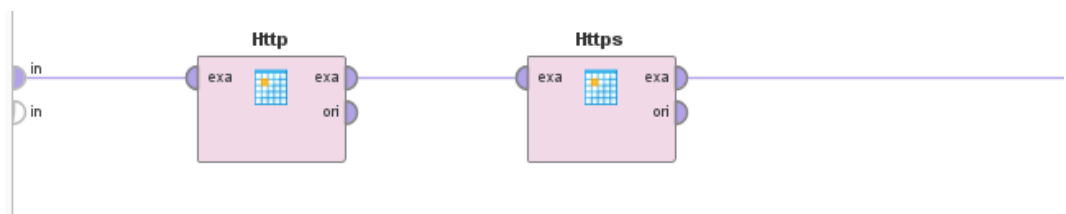


Gambar 4.2.1 Implementasi Algoritma *K-Nearest Neighbor* Rapidminer

Fitur yang digunakan dalam pengimplementasian metode *Klasifikasi K-Nearest Neighbor* ada beberapa yaitu diantaranya :

4.2.1 Fitur Subprocess

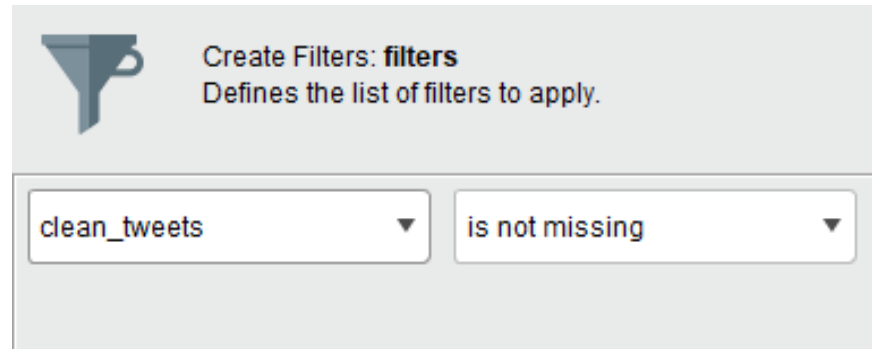
Fitur *Subprocess* pada penelitian ini digunakan untuk menampung fitur “*Replace*” yang mana akan digunakan untuk menghapus kata-kata yang tidak penting seperti *link url* yang tidak bisa dianggap sebagai suatu penilaian dalam menentukan klasifikasi data analisis sentiment pada penelitian ini.



Gambar 4.2.1.1 Fitur *Subprocess* dan *Replace*

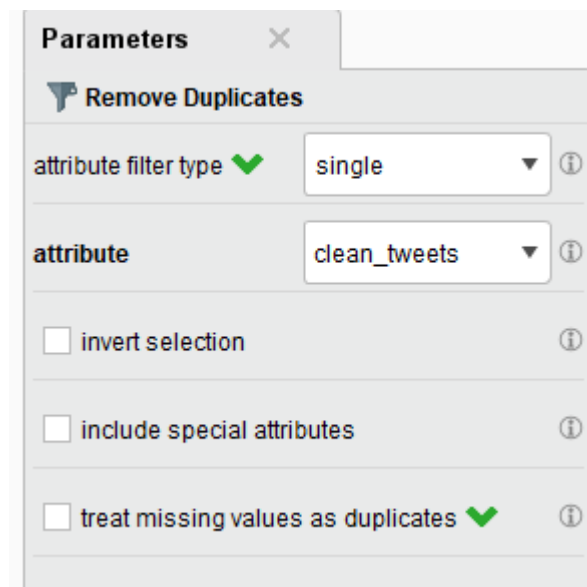
4.2.2 Fitur Filter Example dan Remove Duplicate

Fitur *filter example* digunakan untuk menentukan kondisi yang sesuai dengan kebutuhan penelitian. Kondisi yang digunakan pada penelitian kali ini adalah dimana variable “clean_tweets” tidak boleh ada yang hilang atribut atau data.



Gambar 4.2.2.1 Fitur *Filter Example*

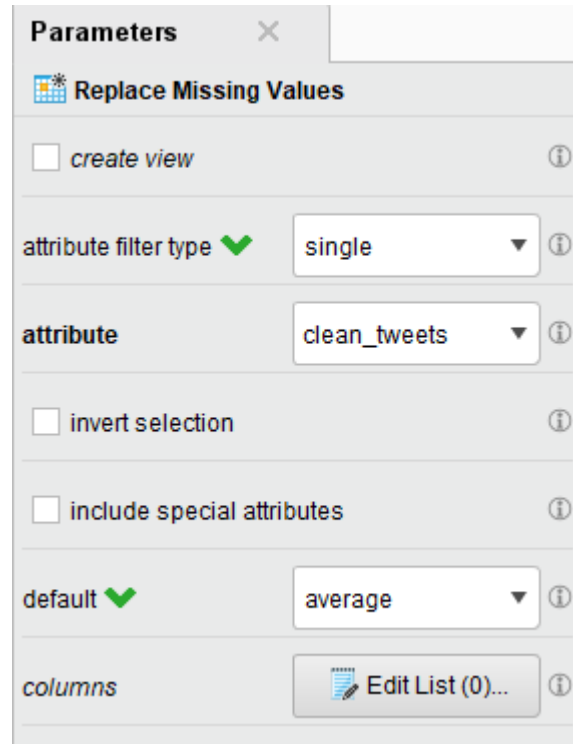
Sedangkan untuk fitur *Remove Duplicate* digunakan untuk menghapus data yang muncul lebih dari satu kali. Variable yang digunakan dalam menentukan data mana yang tidak boleh ada yang sama adalah “clean_tweets”.



Gambar 4.2.2.2 Fitur *Remove Duplicate*

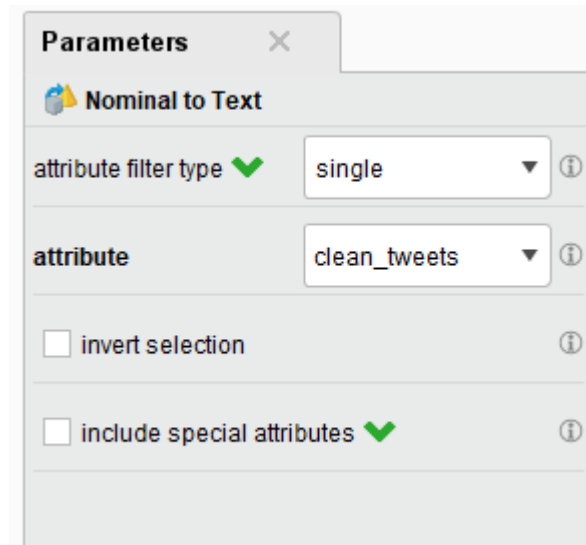
4.2.3 Fitur Replace Missing Value dan Nominal to Text

Fitur *Replace Missing Value* akan menggantikan nilai dari atribut yang hilang dalam data yang akan digunakan. Parameter yang digunakan adalah *single* untuk atribut “clean_tweets”.



Gambar 4.2.3.1 Fitur *Remove Duplicate*

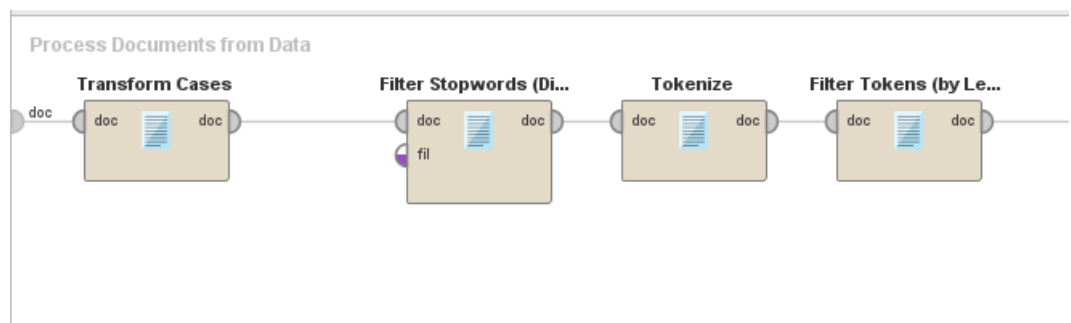
Fitur *Nominal to Text* merupakan fungsi yang digunakan untuk mengubah kata benda menjadi text sehingga bisa dibaca menjadi nilai yang patut diperhatikan nilainya. Parameter yang dipakai pada fitur ini adalah *single* untuk atribut “clean_tweets”.



Gambar 4.2.3.2 Fitur *Nominal to Text*

4.2.4 Fitur Process Documents from Data

Fitur ini digunakan untuk memproses dokumen dari data yang sudah diinput dan diolah untuk mengetahui nilai dari bobot setiap kata-katanya seberapa sering kata itu muncul.



Gambar 4.2.4.1 Fitur Process Documents from Data

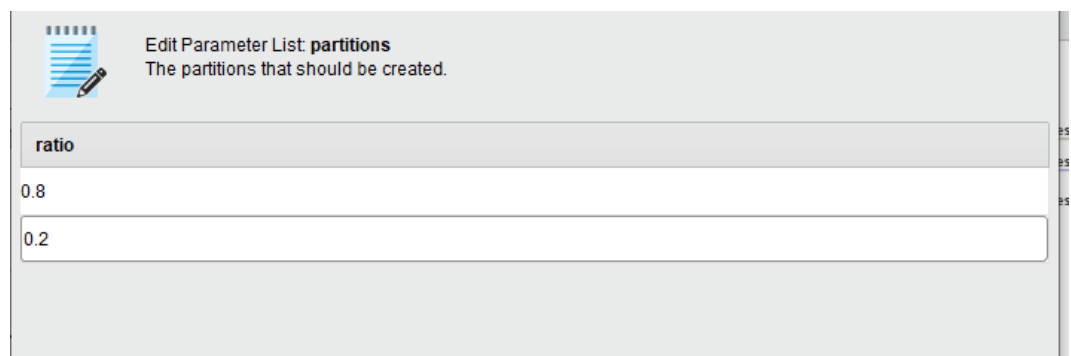
Masing-masing memiliki fitur memiliki kegunaan sebagai berikut :

- c. *Transform Cases* digunakan untuk mengubah semua huruf menjadi huruf kecil semua
- d. *Filter Stopwords* merupakan perintah yang digunakan untuk menyaring kata-kata yang tidak akan dipakai
- e. *Tokenize* digunakan untuk memecah kalimat menjadi kata-kata yang dipisahkan.

- f. *Filter Tokenize* merupakan filter kata yang akan dinilai dari seberapa Panjang kata tersebut, disini parameter yang digunakan adalah “Min Char : 4” dan “Max Char : 50”

4.2.5 Fitur *Split Data*

Fitur ini digunakan untuk membagi data sesuai dengan kebutuhan dengan penelitian. Pada kali ini data akan diberi rasio menjadi 80% : 20% atau 0,8 : 0,2 maka parameter yang digunakan sebagai berikut :

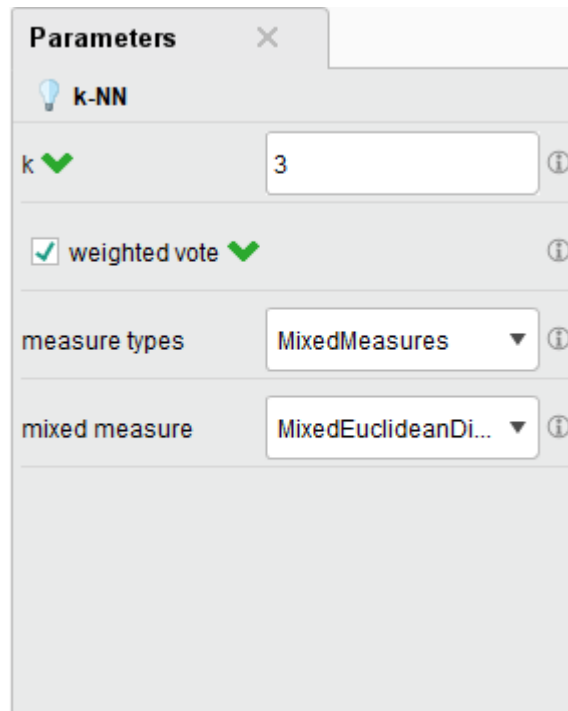


Gambar 4.2.5.1 Fitur *Split Data*

Pada gambar 4.2.1 merupakan model yang akan digunakan dalam menentukan prediksi pada label data yang sudah melewati tahapan training untuk melihat seberapa akurat data yang akan digunakan pada model *Klasifikasi K-Nearest Neighbor*. Untuk melihat hasil dari akurasi maka “K” yang akan digunakan adalah K=3, K=5, K=7 dan K=9.

Tabel 4.2.1 Parameter data yang digunakan

Parameter “K”	3, 5, 7, 9
Measure Types	MixedMeasures
Mixed Measures	MixedEcludianDistance



Gambar 4.2.2 Atribut Algoritma *K-Nearest Neighbor* Rapidminer

Pada percobaan pertama dengan parameter $K=3$ didapatkan hasil sebagai berikut :

accuracy: 56.74%

	true Positive	true Negative	true Neutral	class precision
pred. Positive	14	8	3	56.00%
pred. Negative	7	20	7	58.82%
pred. Neutral	22	14	46	56.10%
class recall	32.56%	47.62%	82.14%	

Gambar 4.2.3 Akurasi Prediksi Percobaan Pertama

Pada percobaan pertama dengan parameter $K=5$ didapatkan hasil sebagai berikut :

accuracy: 52.48%

	true Positive	true Negative	true Neutral	class precision
pred. Positive	9	6	5	45.00%
pred. Negative	8	21	7	56.33%
pred. Neutral	26	15	44	51.76%
class recall	20.93%	50.00%	78.57%	

Gambar 4.2.4 Akurasi Prediksi Percobaan Kedua.

Pada percobaan pertama dengan parameter $K=7$ didapatkan hasil sebagai berikut :

accuracy: 56.03%

	true Positive	true Negative	true Neutral	class precision
pred. Positive	11	4	5	55.00%
pred. Negative	9	25	8	59.52%
pred. Neutral	23	13	43	54.43%
class recall	25.56%	59.52%	76.79%	

Gambar 4.2.5 Akurasi Prediksi Percobaan Ketiga

Pada percobaan pertama dengan parameter $K=9$ didapatkan hasil sebagai berikut :

accuracy: 53.19%

	true Positive	true Negative	true Neutral	class precision
pred. Positive	11	4	7	50.00%
pred. Negative	12	24	9	53.33%
pred. Neutral	20	14	40	54.06%
class recall	25.56%	57.14%	71.43%	

Gambar 4.2.6 Akurasi Prediksi Percobaan Keempat

4.3 Evaluasi Performance

Dari percobaan pertama sampai keempat didapatkan hasil sebagai berikut :

Tabel 4.3.1 Akurasi “K” Performance

Jumlah “K”	Tingkat Akurasi
K=3	56,74%
K=5	52,48%
K=7	56,03%
K=9	53,19%

Berdasarkan tabel 4.3.1 dapat disimpulkan bahwa jumlah “K” yang memiliki tingkat akurasi paling tinggi adalah $K=3$ dengan tingkat akurasi sebesar 56,74%. Ini merupakan data sebelum dan sesudah diimplementasikan pada model KNN dengan data Training yang sudah dibagi menggunakan Rapidminer :

Tabel 4.3.2 Data Awal dan Data Prediksi dengan “K”

Label	Data Awal	K=3	K=5	K=7	K=9
Positif	18	25	20	20	22
Netral	62	82	85	79	74
Negatif	61	34	36	42	45
Total	141	141	141	141	141

4.4 Visualisasi Data

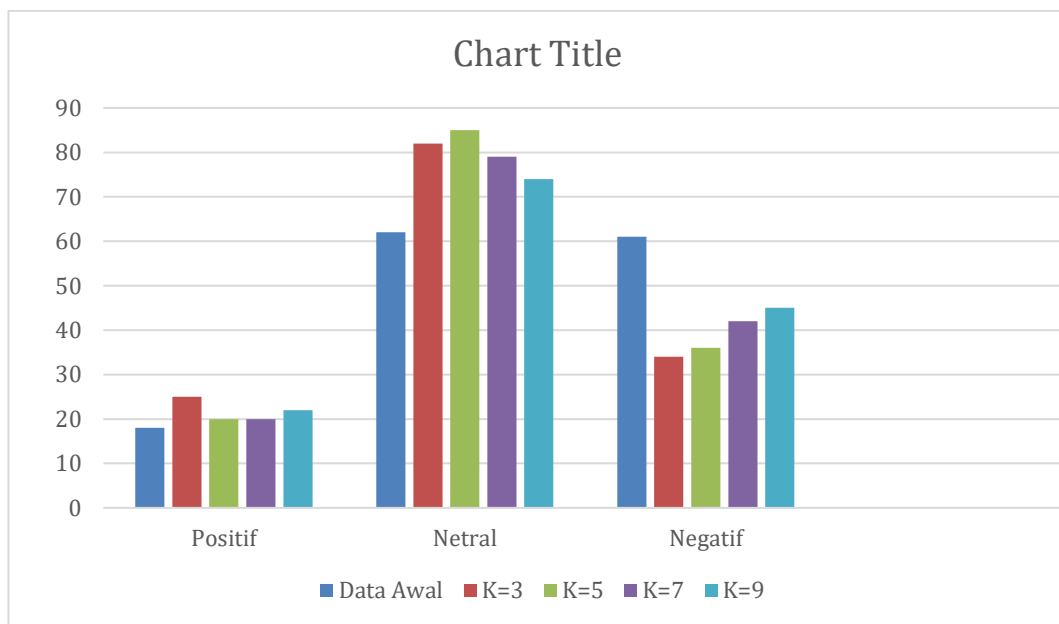
Visualisasi data yang digunakan dalam penelitian ini adalah menggunakan “*Wordlist to Data*”. Yang mana akan diambil dengan tingkat akurasi yang paling tinggi yaitu K=3 sebesar 56,74%.



Gambar 4.4.1 Wordcloud Data Kata Yang Paling Banyak Muncul



Gambar 4.4.1 *Wordcloud* Data Sentimen Paling Banyak Muncul



Gambar 4.4.1 *Chart* Hasil Prediksi

Kelebihan:

- Dalam proses labeling lebih cepat dikarenakan menggunakan *Library Python Textblob*.
- Pengambilan dan pengolahan data lebih cepat dikarenakan menggunakan aplikasi "*Rapidminer*".
- Data yang didapatkan menggunakan Aplikasi "*Rapidminer*" adalah data yang terbaru.

Kekurangan:

- Jika data yang diolah banyak maka dalam proses "*Labeling*" akan berlangsung lama dan akan ada kemungkinan untuk gagal.
- Data perlu diperhatikan dengan seksama agar tidak terdapat data yang sama atau hilang.
- Model yang digunakan dalam penelitian ini kurang dalam memberikan data Training dikarenakan alat tidak mampu dalam melaksanakannya.