

```
pip install Sastrawi
```

```
Looking in indexes: https://pypi.org/simple, https://us-python.pkg.dev/colab-wheels/public/simple/
Collecting Sastrawi
  Downloading Sastrawi-1.0.1-py2.py3-none-any.whl (209 kB)
    _____ 209.7/209.7 kB 14.1 MB/s eta 0:00:00
Installing collected packages: Sastrawi
Successfully installed Sastrawi-1.0.1
```

```
!pip install -U scikit-learn
```

```
Looking in indexes: https://pypi.org/simple, https://us-python.pkg.dev/colab-wheels/public/simple/
Requirement already satisfied: scikit-learn in /usr/local/lib/python3.10/dist-packages (1.2.2)
Requirement already satisfied: numpy>=1.17.3 in /usr/local/lib/python3.10/dist-packages (from scikit-learn) (1.22.4)
Requirement already satisfied: scipy>=1.3.2 in /usr/local/lib/python3.10/dist-packages (from scikit-learn) (1.10.1)
Requirement already satisfied: joblib>=1.1.1 in /usr/local/lib/python3.10/dist-packages (from scikit-learn) (1.2.0)
Requirement already satisfied: threadpoolctl>=2.0.0 in /usr/local/lib/python3.10/dist-packages (from scikit-learn) (3.1.0)
```

```
#LIBRARY
```

```
import pandas as pd
pd.set_option('display.max_colwidth', -1)
from wordcloud import WordCloud
import re
import string
from wordcloud import WordCloud
```

```
from nltk.tokenize import word_tokenize
from string import punctuation
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
sns.set(style="darkgrid")
sns.set(font_scale=1.3)
```

```
from sklearn.feature_extraction.text import CountVectorizer
from sklearn.feature_extraction.text import TfidfVectorizer
```

```
from Sastrawi.Stemmer.StemmerFactory import StemmerFactory
factory = StemmerFactory()
stemmer = factory.create_stemmer()
```

```
<ipython-input-3-f9bca1d27dad>:3: FutureWarning: Passing a negative integer is deprecated in version 1.0 and will not be supported in fu
pd.set_option('display.max_colwidth', -1)
```

```
#IMPORT DATA
```

```
from google.colab import drive
drive.mount("/content/gdrive", force_remount=True)
```

```
data = ('/content/gdrive/MyDrive/Dataset/dataset.csv')
df = pd.read_csv(data)
df
```

Mounted at /content/gdrive

	No.	Nama Instagram	Komentar	Kategori	Tanggal Posting	Nama Akun IG Artis/Selebgram	Unnamed: 6	Unnamed: 7
0	1	@delliananda	"Kaka tidur yaa, udah pagi, gaboleh capek2"	Non-bullying	14 Oktober 2019	@isyanasarasvati	NaN	NaN
1	2	@fenninbl	"makan nasi padang aja begini badannya"	Non-bullying	14 Oktober 2019	@isyanasarasvati	NaN	NaN
2	3	@abdurahmanshq	"yang aku suka dari dia adalah selalu cukur jembut sebelum manggung"	Bullying	14 Oktober 2019	@isyanasarasvati	NaN	NaN
3	4	@naila.yoo	"Hai kak Isyana aku ngefans banget sama kak Isyana.aku paling suka lagu kak Isyana itu lagu tetap didalam jiwa"	Non-bullying	14 Oktober 2019	@isyanasarasvati	NaN	NaN

#MISSING VALUE
df.drop('Unnamed: 6', axis=1, inplace=True)
df.drop('Unnamed: 7', axis=1, inplace=True)
df

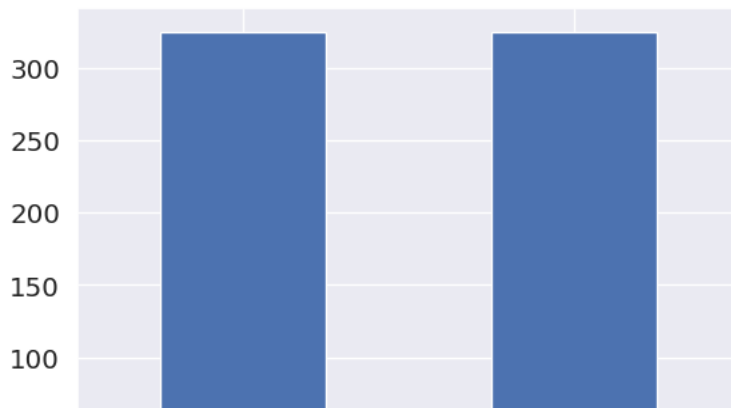
	No.	Nama Instagram	Komentar	Kategori	Tanggal Posting	Nama Akun IG Artis/Selebgram
0	1	@delliananda	"Kaka tidur yaa, udah pagi, gaboleh capek2"	Non-bullying	14 Oktober 2019	@isyanasarasvati
1	2	@fenninbl	"makan nasi padang aja begini badannya"	Non-bullying	14 Oktober 2019	@isyanasarasvati
2	3	@abdurahmanshq	"yang aku suka dari dia adalah selalu cukur jembut sebelum manggung"	Bullying	14 Oktober 2019	@isyanasarasvati
3	4	@naila.yoo	"Hai kak Isyana aku ngefans banget sama kak Isyana.aku paling suka lagu kak Isyana itu lagu tetap didalam jiwa"	Non-bullying	14 Oktober 2019	@isyanasarasvati
4	5	@dessy_____	"Manusia apa bidadari sih herann deh cantik terus 😊♥"	Non-bullying	14 Oktober 2019	@isyanasarasvati
...
645	646	@_sigesrek	"aku memutuskan untuk menjadi fans isyana. gila keren banget ya, otak mereka berdua bisa berfungsi dg menjalankan perintah ke tangan dan kaki, di waktu bersamaan"	Non-bullying	13 Februari 2021	@isyanasarasvati
646	647	@safronlux.id	"AMZING ISYANAAA!! Jujur aku amazed banget dengan skill dan minat Isyana dalam bermusik. Ga Cuma bernyanyi tapi main alat musik pun jago banget"	Non-bullying	13 Februari 2021	@isyanasarasvati
647	648	@rikzikmuktyana	"paling ngiri liat orang keren maen alat musik dan dalem banget nikmatin permainan sendiri"	Non-bullying	13 Februari 2021	@isyanasarasvati
648	649	@antoniusbennys	"Sampe ga bisa berkata2 lagi buat isyana, sangat2 sempurna org ini"	Non-bullying	13 Februari 2021	@isyanasarasvati
649	650	@dikha.wirasathya	"Inimah bukan main alat musik lagi. Olahraga jari dan kaki ini mah"	Non-bullying	13 Februari 2021	@isyanasarasvati

650 rows × 6 columns

```
#GRAFIK NILAI LABEL
import matplotlib.pyplot as plt
%matplotlib inline

df['Kategori'].value_counts().plot.bar()
plt.show()

df['Kategori'].value_counts()
```



```
data = pd.read_csv("/content/gdrive/My Drive/Dataset/dataset.csv")
data = data[["Komentar", "Kategori"]]
data.head

<bound method NDFrame.head of
Komentar \
0  "Kaka tidur yaa, udah pagi, gaboleh capek2"
1  "makan nasi padang aja begini badannya"
2  "yang aku suka dari dia adalah selalu cukur jembut sebelum manggung"
3  "Hai kak Isyana aku ngefans banget sama kak Isyana.aku paling suka lagu kak Isyana itu lagu tetap didalam jiwa"
4  "Manusia apa bidadari sih herann deh cantik terus 😊❤️"
..
645 "aku memutuskan untuk menjadi fans isyana. gila keren banget ya, otak mereka berdua bisa berfungsi dg menjalankan perintah ke
tangan dan kaki, di waktu bersamaan"
646 "AMZING ISYANAA!! Jujur aku amazed banget dengan skill dan minat Isyana dalam bermusik. Ga Cuma bernyanyi tapi main alat musik pun
jago banget"
647 "paling ngiri liat orang keren maen alat musik dan dalem banget nikmatin permainan sendiri"
648 "Sampe ga bisa berkata2 lagi buat isyana, sangat2 sempurna org ini"
649 "Inimah bukan main alat musik lagi. Olahraga jari dan kaki ini mah"

Kategori
0  Non-bullying
1  Non-bullying
2  Bullying
3  Non-bullying
4  Non-bullying
..
645 Non-bullying
646 Non-bullying
647 Non-bullying
648 Non-bullying
649 Non-bullying

[650 rows x 2 columns]>
```

▼ PRE-PROCESSING

```
#CLEANING
def cleaning(text):
    #menghapus tab, new line, dan backslash
    text=text.replace('\t', " ").replace('\n', " ").replace('\u', " ").replace('\\', " ")
    #menghapus non ASCII (emoji, chinese words, etc)
    text=text.encode('ascii', 'replace').decode('ascii')
    #menghapus mention, link, hashtag
    text=" ".join(re.sub("([@#][A-Za-z0-9]+)|(\w+:\w+\/\S+)", " ", text).split())
    #menghapus incomplete URL
    text=text.replace("http://", " ").replace("https://", " ")
    #menghapus angka
    text=re.sub(r"\d+", "",text)
    #menghapus tanda baca
    text=text.translate(str.maketrans("", "",string.punctuation))
    #menghapus spasi di depan dan di belakang
    text=text.strip()
    #menghapus double atau single spasi
    text=re.sub('\s+', ' ',text)
    #menghapus karakter tunggal
    return re.sub(r"\b[a-zA-z]\b", " ", text)
```

```
data['Cleaning'] = data['Komentar'].apply(cleaning)
data
```

	Komentar	Kategori	Cleaning
0	"Kaka tidur yaa, udah pagi, gaboleh capek2"	Non-bullying	Kaka tidur yaa udah pagi gaboleh capek
1	"makan nasi padang aja begini badannya"	Non-bullying	makan nasi padang aja begini badannya
2	"yang aku suka dari dia adalah selalu cukur jembut sebelum manggung"	Bullying	yang aku suka dari dia adalah selalu cukur jembut sebelum manggung
3	"Hai kak Isyana aku ngefans banget sama kak Isyana.aku paling suka lagu kak Isyana itu lagu tetap didalam jiwa"	Non-bullying	Hai kak Isyana aku ngefans banget sama kak Isyanaaku paling suka lagu kak Isyana itu lagu tetap didalam jiwa
4	"Manusia apa bidadari sih herann deh cantik terus 😊♥"	Non-bullying	Manusia apa bidadari sih herann deh cantik terus
...
645	"aku memutuskan untuk menjadi fans isyana. gila keren banget ya, otak mereka berdua bisa berfungsi dg menjalankan perintah ke tangan dan kaki, di waktu bersamaan"	Non-bullying	aku memutuskan untuk menjadi fans isyana gila keren banget ya otak mereka berdua bisa berfungsi dg menjalankan perintah ke tangan dan kaki di waktu bersamaan
646	"AMZING ISYANAA!! Jujur aku amazed banget dengan skill dan minat Isyana dalam bermusik. Ga Cuma bernyanyi tapi main alat musik pun jago banget"	Non-bullying	AMZING ISYANAA Jujur aku amazed banget dengan skill dan minat Isyana dalam bermusik Ga Cuma bernyanyi tapi main alat musik pun jago banget
647	"paling ngiri liat orang keren maen alat musik dan dalem banget nikmatin permainan sendiri"	Non-bullying	paling ngiri liat orang keren maen alat musik dan dalem banget nikmatin permainan sendiri
648	"Sampe ga bisa berkata2 lagi buat isyana, sangat2 sempurna org ini"	Non-bullying	Sampe ga bisa berkata lagi buat isyana sangat sempurna org ini
649	"Inimah bukan main alat musik lagi. Olahraga jari dan kaki ini mah"	Non-bullying	Inimah bukan main alat musik lagi Olahraga jari dan kaki ini mah

650 rows × 3 columns



```
#CASE FOLDING
def casefolding (text):
    return text.lower()

data['CaseFolding'] = data['Cleaning'].apply(casefolding)
data
```

```

#TOKENISASI
def tokenizing(text):
    return word_tokenize(text)

import nltk
nltk.download("punkt")

[nltk_data] Downloading package punkt to /root/nltk_data...
[nltk_data] Unzipping tokenizers/punkt.zip.
True

data['Tokenisasi']=data['CaseFolding'].apply(tokenizing)
data

```

	Komentar	Kategori	Cleaning	CaseFolding	Tokenisasi
0	"Kaka tidur yaa, udah pagi, gaboleh capek2"	Non-bullying	Kaka tidur yaa udah pagi gaboleh capek	kaka tidur yaa udah pagi gaboleh capek	[kaka, tidur, yaa, udah, pagi, gaboleh, capek]
1	"makan nasi padang aja begini badannya"	Non-bullying	makan nasi padang aja begini badannya	makan nasi padang aja begini badannya	[makan, nasi, padang, aja, begini, badannya]
2	"yang aku suka dari dia adalah selalu cukur jembut sebelum manggung"	Bullying	yang aku suka dari dia adalah selalu cukur jembut sebelum manggung	yang aku suka dari dia adalah selalu cukur jembut sebelum manggung	[yang, aku, suka, dari, dia, adalah, selalu, cukur, jembut, sebelum, manggung]
3	"Hai kak Isyana aku ngefans banget sama kak Isyana.aku paling suka lagu kak Isyana itu lagu tetap didalam jiwa"	Non-bullying	Hai kak Isyana aku ngefans banget sama kak Isyanaaku paling suka lagu kak Isyana itu lagu tetap didalam jiwa	hai kak isyana aku ngefans banget sama kak isyanaaku paling suka lagu kak isyana itu lagu tetap didalam jiwa	[hai, kak, isyana, aku, ngefans, banget, sama, kak, isyanaaku, paling, suka, lagu, kak, isyana, itu, lagu, tetap, didalam, jiwa]
4	"Manusia apa bidadari sih herann deh cantik terus 😊♥"	Non-bullying	Manusia apa bidadari sih herann deh cantik terus	manusia apa bidadari sih herann deh cantik terus	[manusia, apa, bidadari, sih, herann, deh, cantik, terus]
...
645	"aku memutuskan untuk menjadi fans isyana. gila keren banget ya, otak mereka berdua bisa berfungsi dg menjalankan perintah ke tangan dan kaki, di waktu bersamaan"	Non-bullying	aku memutuskan untuk menjadi fans isyana gila keren banget ya otak mereka berdua bisa berfungsi dg menjalankan perintah ke tangan dan kaki di waktu bersamaan	aku memutuskan untuk menjadi fans isyana gila keren banget ya otak mereka berdua bisa berfungsi dg menjalankan perintah ke tangan dan kaki di waktu bersamaan	[aku, memutuskan, untuk, menjadi, fans, isyana, gila, keren, banget, ya, otak, mereka, berdua, bisa, berfungsi, dg, menjalankan, perintah, ke, tangan, dan, kaki, di, waktu, bersamaan]
646	"AMZING ISYANAA!! Jujur aku amazed banget dengan skill dan minat Isyana dalam bermusik. Ga Cuma bernyanyi tapi main alat musik pun jago banget"	Non-bullying	AMZING ISYANAA Jujur aku amazed banget dengan skill dan minat Isyana dalam bermusik Ga Cuma bernyanyi tapi main alat musik pun jago banget	amzing isyanaa jujur aku amazed banget dengan skill dan minat isyana dalam bermusik ga cuma bernyanyi tapi main alat musik pun jago banget	[amzing, isyanaa, jujur, aku, amazed, banget, dengan, skill, dan, minat, isyana, dalam, bermusik, ga, cuma, bernyanyi, tapi, main, alat, musik, pun, jago, banget]
647	"paling ngiri liat orang keren maen alat musik dan dalem banget nikmatin permainan sendiri"	Non-bullying	paling ngiri liat orang keren maen alat musik dan dalem banget nikmatin permainan sendiri	paling ngiri liat orang keren maen alat musik dan dalem banget nikmatin permainan sendiri	[paling, ngiri, liat, orang, keren, maen, alat, musik, dan, dalem, banget, nikmatin, permainan, sendiri]
648	"Sampe ga bisa berkata2 lagi buat isyana, sangat2 sempurna org ini"	Non-bullying	Sampe ga bisa berkata lagi buat isyana sangat sempurna org ini	sampe ga bisa berkata lagi buat isyana sangat sempurna org ini	[sampe, ga, bisa, berkata, lagi, buat, isyana, sangat, sempurna, org, ini]

```

#NORMALISASI
key_norm = pd.read_csv('/content/gdrive/My Drive/Dataset/gabungan.csv')
def normalizing(text):
    text=' '.join([key_norm[key_norm['awal'] == word][ 'akhir'].values[0] if (key_norm["awal"] == word).any() else word for word in text])
    return text.split(' ')

data['Normalisasi']=data['Tokenisasi'].apply(normalizing)
data

```

	Komentar	Kategori	Cleaning	CaseFolding	Tokenisasi	Normalisasi
0	"Kaka tidur yaa, udah pagi, gaboleh capek2"	Non-bullying	Kaka tidur yaa udah pagi gaboleh capek	kaka tidur yaa udah pagi gaboleh capek	[kaka, tidur, yaa, udah, pagi, gaboleh, capek]	[kakak, tidur, ya, sudah, pagi, tidak, boleh, capai]
1	"makan nasi padang aja begini badannya"	Non-bullying	makan nasi padang aja begini badannya	makan nasi padang aja begini badannya	[makan, nasi, padang, aja, begini, badannya]	[makan, nasi, padang, saja, begini, badannya]
2	"yang aku suka dari dia adalah selalu cukur jembut sebelum mengganggu"	Bullying	yang aku suka dari dia adalah selalu cukur jembut sebelum mengganggu	yang aku suka dari dia adalah selalu cukur jembut sebelum mengganggu	[yang, aku, suka, dari, dia, adalah, selalu, cukur, jembut, sebelum, mengganggu]	[yang, aku, suka, dari, dia, adalah, selalu, cukur, jembut, sebelum, panggung]
3	"Hai kak Isyana aku ngefans banget sama kak Isyana.aku paling suka lagu kak Isyana itu lagu tetap didalam jiwa"	Non-bullying	Hai kak Isyana aku ngefans banget sama kak Isyanaaku paling suka lagu kak Isyana itu lagu tetap didalam jiwa	hai kak isyana aku ngefans banget sama kak isyanaaku paling suka lagu kak isyana itu lagu tetap didalam jiwa	[hai, kak, isyana, aku, ngefans, banget, sama, kak, isyanaaku, paling, suka, lagu, kak, isyana, itu, lagu, tetap, didalam, jiwa]	[hai, kakak, isyana, aku, ngefans, sekali, sama, kakak, isyana, aku, paling, suka, lagu, kakak, isyana, itu, lagu, tetap, didalam, jiwa]
4	"Manusia apa bidadari sih herann deh cantik terus 😊♥"	Non-bullying	Manusia apa bidadari sih herann deh cantik terus	manusia apa bidadari sih herann deh cantik terus	[manusia, apa, bidadari, sih, herann, deh, cantik, terus]	[manusia, apa, bidadari, sih, heran, deh, cantik, terus]
...
645	"aku memutuskan untuk menjadi fans isyana. gila keren banget ya, otak mereka berdua bisa berfungsi dg menjalankan"	Non-bullying	aku memutuskan untuk menjadi fans isyana gila keren banget ya otak mereka berdua bisa berfungsi dg menjalankan	aku memutuskan untuk menjadi fans isyana gila keren banget ya otak mereka berdua bisa berfungsi dg menjalankan	[aku, memutuskan, untuk, menjadi, fans, isyana, gila, keren, banget, ya, otak, mereka, berdua, bisa, berfungsi, dg, menjalankan,	[aku, memutuskan, untuk, menjadi, fans, isyana, gila, keren, sekali, ya, otak, mereka, berdua, bisa, berfungsi, dengan,

```

#STOPWORDS REMOVAL
stopwords = pd.read_csv('/content/gdrive/My Drive/Dataset/stopwords_baru.csv')
extend = pd.DataFrame(['yg','dg','rt','dgn','ny','d','klo','kalo','amp','bier','bikin','bilang'
    'gak','ga','krn','nya','nih','sih','si','tau','tdk','tuh','utk','ya','jd','jgn',
    'sdh','aja','n','t','nyg','hehe','pen','u','wkwk','nan','loh','&','yah','my',
    'rb','jr','rp','hr','di','kb','gb'])
stopwords=stopwords.append(extend, ignore_index=True)

list_stopwords = set(stopwords.iloc[:,0])

def stopword_removing(words):
    return [word for word in words if word not in list_stopwords]

<ipython-input-17-486808f01a61>:7: FutureWarning: The frame.append method is deprecated and will be removed from pandas in a future vers
    stopwords=stopwords.append(extend, ignore_index=True)

data['StopwordRemoval'] = data['Normalisasi'].apply(stopword_removing)
data

```

	Komentar	Kategori	Cleaning	CaseFolding	Tokenisasi	Normalisasi	StopwordRemoval
0	"Kaka tidur yaa, udah pagi, gaboleh capek2"	Non-bullying	Kaka tidur yaa udah pagi gaboleh capek	kaka tidur yaa udah pagi gaboleh capek	[kaka, tidur, yaa, udah, pagi, gaboleh, capek]	[kakak, tidur, ya, sudah, pagi, tidak, boleh, capai]	[kakak, tidur, pagi, capai]
1	"makan nasi padang aja begini badannya"	Non-bullying	makan nasi padang aja begini badannya	makan nasi padang aja begini badannya	[makan, nasi, padang, aja, begini, badannya]	[makan, nasi, padang, saja, begini, badannya]	[makan, nasi, padang, badannya]
2	"yang aku suka dari dia adalah selalu cukur jembut sebelum manggung"	Bullying	yang aku suka dari dia adalah selalu cukur jembut sebelum manggung	yang aku suka dari dia adalah selalu cukur jembut sebelum manggung	[yang, aku, suka, dari, dia, adalah, selalu, cukur, jembut, sebelum, manggung]	[yang, aku, suka, dari, dia, adalah, selalu, cukur, jembut, sebelum, manggung]	[suka, cukur, jembut, panggung]
3	"Hai kak Isyana aku ngefans banget sama kak Isyana.aku paling"	Non-	Hai kak Isyana aku ngefans banget sama kak Isyanaaku	hai kak isyana aku ngefans banget sama kak isyanaaku paling	[hai, kak, isyana, aku, ngefans, banget, sama, kak, isyanaaku,]	[hai, kakak, isyana, aku, ngefans, sekali, sama, kakak, isyana,]	[hai, kakak, isyana, kakak, isyana, suka,]
#STEMMING							
def stemming(words):							
stem_kalimat = []							
for k in words:							
stem_kata = stemmer.stem(k)							
stem_kalimat.append(stem_kata)							
stem_kalimat_str = ' '.join(stem_kalimat)							
return stem_kalimat_str							
data['Stemming']=data['StopwordRemoval'].apply(stemming)							
data							

	Komentar	Kategori	Cleaning	CaseFolding	Tokenisasi	Normalisasi	StopwordRemoval	Stemming
0	"Kaka tidur yaa, udah pagi, gaboleh capek2"	Non-bullying	Kaka tidur yaa udah pagi gaboleh capek	kaka tidur yaa udah pagi gaboleh capek	[kaka, tidur, yaa, udah, pagi, gaboleh, capek]	[kakak, tidur, ya, sudah, pagi, tidak, boleh, capai]	[kakak, tidur, pagi, capai]	kakak tidur pagi capai
1	"makan nasi padang aja begini badannya"	Non-bullying	makan nasi padang aja begini badannya	makan nasi padang aja begini badannya	[makan, nasi, padang, aja, begini, badannya]	[makan, nasi, padang, saja, begini, badannya]	[makan, nasi, padang, badannya]	makan nasi padang badan
2	"yang aku suka dari dia adalah selalu cukur jembut sebelum manggung"	Bullying	yang aku suka dari dia adalah selalu cukur jembut sebelum manggung	yang aku suka dari dia adalah selalu cukur jembut sebelum manggung	[yang, aku, suka, dari, dia, adalah, selalu, cukur, jembut, sebelum, manggung]	[yang, aku, suka, dari, dia, adalah, selalu, cukur, jembut, sebelum, manggung]	[suka, cukur, jembut, panggung]	suka cukur jembut panggung
3	"Hai kak Isyana aku ngefans banget sama kak Isyana.aku paling suka lagu kak Isyana itu lagu tetap didalam jiwa"	Non-bullying	Hai kak Isyana aku ngefans banget sama kak Isyanaaku paling suka lagu kak Isyana itu lagu tetap didalam jiwa	hai kak isyana aku ngefans banget sama kak isyanaaku paling suka lagu kak isyana itu lagu tetap didalam jiwa	[hai, kak, isyana, aku, ngefans, banget, sama, kak, isyanaaku, paling, suka, lagu, kak, isyana, itu, lagu, tetap, didalam, jiwa]	[hai, kakak, isyana, aku, ngefans, sekali, sama, kakak, isyana, aku, paling, suka, lagu, kakak, isyana, itu, lagu, tetap, didalam, jiwa]	[hai, kakak, isyana, kakak, isyana, suka, lagu, kakak, isyana, lagu, didalam, jiwa]	hai kakak isyana kakak isyana suka lagu kakak isyana lagu dalam jiwa
4	"Manusia apa bidadari sih herann deh cantik terus 🥰"	Non-bullying	Manusia apa bidadari sih herann deh cantik terus	manusia apa bidadari sih herann deh cantik terus	[manusia, apa, bidadari, sih, herann, deh, cantik, terus]	[manusia, apa, bidadari, sih, heran, deh, cantik, terus]	[manusia, bidadari, heran, cantik]	manusia bidadari heran cantik
...
645	"aku memutuskan untuk menjadi fans isyana. gila keren banget ya, otak mereka berdua bisa berfungsi dg menjalankan perintah ke tangan dan kaki, di waktu bersamaan"	Non-bullying	aku memutuskan untuk menjadi fans isyana gila keren banget ya otak mereka berdua bisa berfungsi dg menjalankan perintah ke tangan dan kaki di waktu bersamaan	aku memutuskan untuk menjadi fans isyana gila keren banget ya otak mereka berdua bisa berfungsi dg menjalankan perintah ke tangan dan kaki di waktu bersamaan	[aku, memutuskan, untuk, menjadi, fans, isyana, gila, keren, banget, ya, otak, mereka, berdua, bisa, berfungsi, dg, menjalankan, perintah, ke, tangan, dan, kaki, di, waktu, bersamaan]	[aku, memutuskan, untuk, menjadi, fans, isyana, gila, keren, sekali, ya, otak, mereka, berdua, bisa, berfungsi, dengan, menjalankan, perintah, ke, tangan, dan, kaki, di, waktu, bersamaan]	[memutuskan, isyana, gila, keren, otak, berdua, berfungsi, menjalankan, perintah, tangan, kaki, bersamaan]	putus isyana gila keren otak dua fungsi jalan perintah tangan kaki sama
646	"AMZING ISYANAA! Jujur aku amazed banget dengan skill dan minat Isyana dalam bermusik. Ga Cuma bernyanyi tapi main alat"	Non-bullying	AMZING ISYANAA Jujur aku amazed banget dengan skill dan minat Isyana dalam bermusik Ga Cuma bernyanyi tapi main alat	amzing isyanaa jujur aku amazed banget dengan skill dan minat isyana dalam bermusik ga cuma bernyanyi tapi main alat	[amzing, isyanaa, jujur, aku, amazed, banget, dengan, skill, dan, minat, isyana, dalam, bermusik, ga, cuma, bernyanyi, tapi, main, alat, musik]	[amzing, isyana, jujur, aku, amazed, sekali, dengan, skill, dan, minat, isyana, dalam, bermusik, tidak, cuma, bernyanyi, tetapi, main, alat, musik]	[isyana, jujur, minat, isyana, bermusik, bernyanyi, main, alat, musik, jago]	isyana jujur minat isyana musik nyanyi main alat musik iago

```
# membuat csv hasil preprocessing
data.to_csv('hasil_preprocessing.csv')
```

EKSTRAKSI FITUR TF-IDF

TF-IDF VER 2


```
# Prepro Data
data = pd.read_csv ("/content/hasil_preprocessing.csv", usecols=["Stemming", "Kategori"])

tfidf = TfidfVectorizer(ngram_range=(1,1))
tweet = data['Stemming'].values.astype('U')
tfidf_matrix = tfidf.fit_transform(tweet)
tweet = pd.DataFrame(tfidf_matrix.toarray(), columns=tfidf.get_feature_names_out(), index=tweet)

# Mengambil Total Jumlah Data Dari Matrix
total_data = tfidf_matrix.shape[0]

# Menyimpan Hasil
tweet.to_csv("tfidf.csv", index=False)

tweet.head()
```



	abai	abal	abang	abu	ada	adab	adam	adaptasi	adik	agama	...	waras	wa
kakak tidur pagi capai	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	
makan nasi padang badan	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	
suka cukur jembut panggung	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	