

# KLASIFIKASI CYBERBULLYING PADA TWITTER

Vicha Farazahla  
2100018400





# Dataset

Dataset didapatkan dari Kaggle, dengan Studi Kasus Twitter Cyberbullying Classification.  
“cyberbullying\_tweets.csv”

## Cyberbullying Classification | 91% + EDA

Python · [Cyberbullying Classification](#)

Notebook Input Output Logs Comments (0)

### Input Data

cyberbullying\_tweets.csv (7.17 MB)

Detail Compact Column

2 of 2 columns

#### About this file

Dataset file from: J. Wang, K. Fu, C.T. Lu, “SOSNet: A Graph Convolutional Network Approach to Fine-Grained Cyberbullying Detection,” Proceedings of the 2020 IEEE International Conference on Big Data (IEEE BigData 2020), December 10-13, 2020.

#### Input (7.17 MB)

Data Sources

Cyberbullying Classific  
cyberbullying\_tweet

	tweet_text	cyberbullying_type
1	In other words #katandandre, your food was crapilicious! #mkr,not_cyberbullying	
2	Why is #aussietv so white? #MKR #theblock #imACelebrityAU #today #sunrise #studio10 #Neighbours #WonderlandTen #etc,not_cyberbullying	
3	@XochitlSuckkks a classy whore? Or more red velvet cupcakes?,not_cyberbullying	
4	@Jason_Gio meh. :P thanks for the heads up, but not too concerned about another angry dude on twitter.,not_cyberbullying	
5	@RudhoeEnglish This is an ISIS account pretending to be a Kurdish account. Like Islam, it is all lies.,not_cyberbullying	
6	@RajaSaab @QuickieLeaks Yes, the test of god is that good or bad or indifferent or weird or whatever, it all proves gods existence.,not_cyberbullying	
7	Itu sekolah ya bukan tempat bully! Ga jauh kaya neraka,not_cyberbullying	
8	Karma. I hope it bites Kat on the butt. She is just nasty. #mkr,not_cyberbullying	
9	@stockputout everything but mostly my priest,not_cyberbullying	
10	Rebecca Black Drops Out of School Due to Bullying,not_cyberbullying	
11	@Jord_is_Dead http://t.co/UsQinYW5Gn,not_cyberbullying	
12	The Bully flushes on KD http://twitvid.com/AZTNP,not_cyberbullying	
13	Ughhhh #MKR,not_cyberbullying	
14	RT @Kurdnews: Turkish state has killed 241 children in last 11 years http://t.co/JlvKE1epws #news ##GoogleÃteviriciTopluluÃirtÃeyideEÃ,	
15	Love that the best response to the hotcakes they managed to film was a non-committal "meh" from some adolescent. #MKR,not_cyberbullying	
16	@yasmimcaci @Berrarii PAREM DE FAZER BULLYING COMIGO -( UHAHUAH BANDO DE PRETO,not_cyberbullying	
17	@sarinhacoral @Victor_Maggi tadinho de mim , sofrendo bullying viu MIMI',not_cyberbullying	
18	@Oxabad1dea @kelseytheodore2 twitter is basically the angry letters of our generation.,not_cyberbullying	
19	Best pick up line? Hi, you're cute... ? : I love how people call James Potter is a bully. - mypatronusisyou: http://tumblr.com/xol3xl14zy,not_cyberbullying	
20	Now I gotta walk to classss?! I officially hate the stupid bus system! -.,not_cyberbullying	
21	@halalcunty @biebervalue @liamkiwi @greenlinerjzm @halalflaws I know Saudis chased girls into a burning building. http://t.co/8b1AclczN9,not_cyberbullying	
22	Kids LoveðŸ™ @ Mohamad Bin Zayed City Ù¸ÙŠÙÙ Ù¸Ù¸ ð¸¸+ ð¸¸ð¸ÙŠð¸ http://t.co/0xrOZSNn,not_cyberbullying	
23	I still have Jack, Amsterdam, Ciroc, Crown, Bud Light, Lime-a-rita, and CRU in my love handles. I feel the jiggle when I walk...,not_cyberbullying	
24	@scottyswaggod men are the ones that are going to push the real change. they are the ones in power, generally. we need their help.,not_cyberbullying	
25	Michine... not_cyberbullying	
26	Michine... not_cyberbullying	

Terdapat 2 atribut dalam dataset “cyberbullying\_tweets.csv” yaitu ‘tweet\_text’ dan ‘cyberbullying\_type’

# Library Python yang Digunakan



Library numpy, pandas,  
pickle, seaborn, matplotlib

```
[3] import numpy as np
import pandas as pd
import pickle
import seaborn as sns
from wordcloud import WordCloud
import matplotlib.pyplot as plt
%matplotlib inline
```

Numpy untuk operasi numerik  
Pandas untuk operasi pada  
tabel data.

Pickle untuk menyimpan model  
yang telah dilatih sehingga  
bisa digunakan kembali tanpa  
harus melatih ulang.

Seaborn dan Matplotlib untuk  
visualisasi data, seperti  
membuat grafik distribusi data  
atau hasil klasifikasi

Library NLTK

```
import nltk
from nltk.stem import WordNetLemmatizer
from nltk.corpus import stopwords
from nltk.tokenize import word_tokenize
from string import punctuation
```

NLTK digunakan untuk berbagai tugas  
pemrosesan teks, seperti tokenisasi,  
penghapusan stopwords, dan  
lemmatization.

Library Sklearn

```
from sklearn.model_selection import train_test_split
from sklearn.feature_extraction.text import TfidfVectorizer
from sklearn.metrics import confusion_matrix, classification_report
from sklearn.linear_model import LogisticRegression
from sklearn.naive_bayes import MultinomialNB
from sklearn.ensemble import RandomForestClassifier
```

sklearn digunakan untuk membagi dataset, mengubah teks  
menjadi fitur numerik, melatih model klasifikasi, dan  
mengevaluasi kinerjanya.

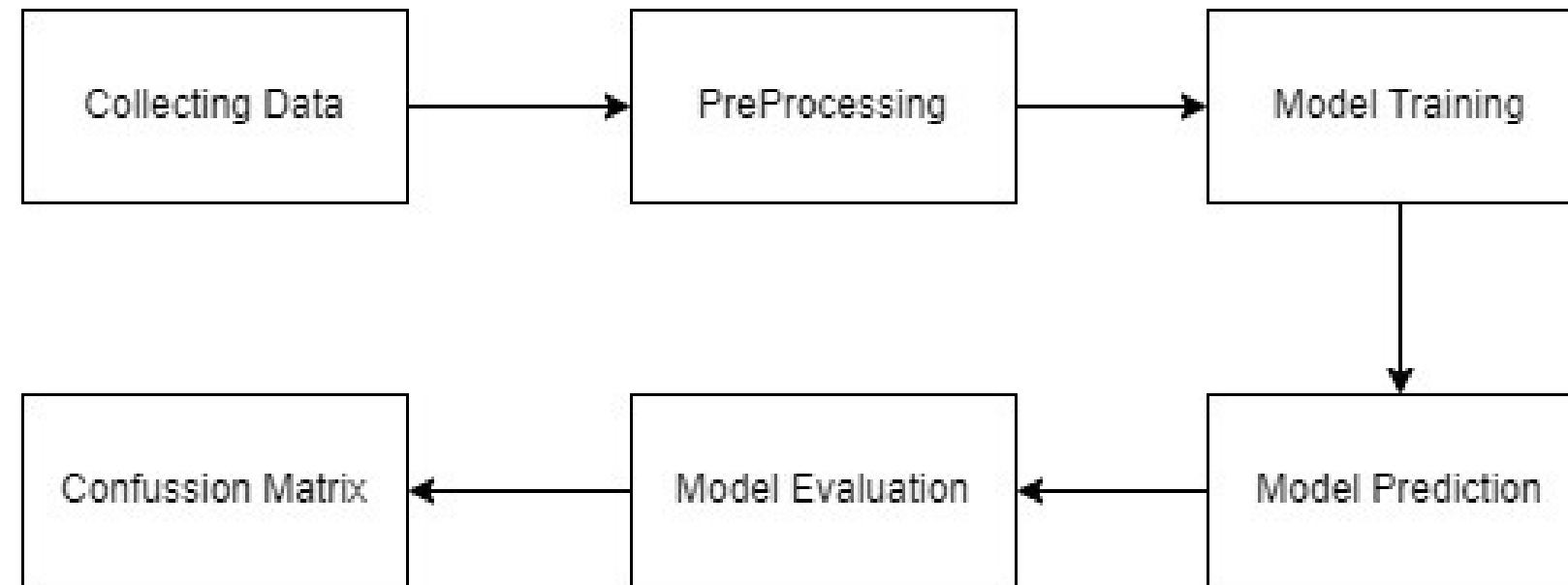


# Tahapan Pembuatan Program

01

## Collecting Data

Mengumpulkan data dengan cara crawling data dari Twitter yang disimpan dalam file "crawlingdt.csv"



02

## PreProcessing

Tahap dilakukan cleaning data, stopword, stemming, tokenizing

03

## Model Training

Model menggunakan algoritma pembelajaran untuk menyesuaikan parameter internalnya agar dapat membuat prediksi yang akurat.

04

## Model Prediction

Menggunakan model yang telah dilatih untuk membuat prediksi pada data baru yang belum pernah dilihat model sebelumnya.

05

## Model Evaluation

Model diuji menggunakan data evaluasi atau data uji (test data), yang berbeda dari data pelatihan.

06

## Confussion Matrix


Memberikan gambaran tentang bagaimana model membuat prediksi, memungkinkan kita untuk melihat berapa banyak prediksi yang benar dan salah dalam setiap kategori

# Hasil Akhir

```
▶ model_file = open('Voting.pickle', 'rb')
model = pickle.load(model_file)
model_file.close()

texts = ["I hate you bitch",
        "School shooting. Haha dead.",
        "Treat them like black slaves",
        "I don't feel so good Mr. Potter",
        "christians are the worst"
        ]

predict(model, texts)
```



	text	type
0	I hate you bitch	gender
1	School shooting. Haha dead.	age
2	Treat them like black slaves	ethnicity
3	I don't feel so good Mr. Potter	not_cyberbullying
4	christians are the worst	religion

Hasil dari proses klasifikasi teks menggunakan model yang telah dilatih. Setiap teks dikategorikan ke dalam tipe atau label tertentu berdasarkan kontennya.

# Kesimpulan

Kualitas dan variasi data latih yang baik sangat penting untuk kinerja model. Penggunaan dataset yang terlabel dengan baik memungkinkan model untuk belajar dan mengenali pola yang lebih kompleks dan nuansa dalam teks.



Insight yang diperoleh dari proyek ini memberikan dasar yang kuat untuk pengembangan lebih lanjut dan aplikasi dalam pemantauan konten dan intervensi yang proaktif terhadap cyberbullying..