

# uts-text-mining

November 11, 2023

LINK VIDEO: <https://drive.google.com/file/d/12tRBnlociu2tiJ8LyA5R1GCo0rXYsTEq/view?usp=sharing>

LINK BACKUP: [https://binusianorg-my.sharepoint.com/personal/fanes\\_hanestan\\_binus\\_ac\\_id/Documents/25](https://binusianorg-my.sharepoint.com/personal/fanes_hanestan_binus_ac_id/Documents/25)

#Import Library

```
[1]: import requests
from bs4 import BeautifulSoup
import pandas as pd
from sklearn.model_selection import train_test_split
from sklearn.svm import SVC
from sklearn.ensemble import RandomForestClassifier
from sklearn.metrics import accuracy_score, classification_report
from sklearn.feature_extraction.text import TfidfVectorizer
import string
import re
import nltk
from nltk.corpus import stopwords
from nltk.stem import WordNetLemmatizer
```

#Extract Text from News Link

```
[2]: def class_media(media):
    if media == 'CNN Indonesia':
        return 'detail-text'
    elif media == 'Liputan6':
        return 'read-page--content'
    elif media == 'Kompas':
        return 'read__content'
    elif media == 'CNBC Indonesia':
        return 'detail_wrap'
    else:
        return ''
```

```
[3]: #sesuai di soal, kita harus mengambil link berita dari berbagai sumber

# Total yang diambil adalah:
# CNN = 2 x 3(topik) = 6 berita
# Liputan6 = 2 x 3 = 6 berita
```

```

# Kompas = 2 x 3 = 6 berita
# CNBC Indonesia = 2 x 3 = 6 berita
# SUM(berita) = 24 berita

link = {
    'https://www.cnnindonesia.com/olahraga/20231031111217-170-1018035/
    ↳dua-sisi-jonatan-christie#anchor': {'Media': 'CNN Indonesia', 'Label': '
    ↳Olahraga'},
    'https://www.cnnindonesia.com/olahraga/20231105160145-178-1020246/
    ↳klasemen-liga-voli-korea-usai-megawati-bawa-red-sparks-sikat-ai': {'Media': '
    ↳CNN Indonesia', 'Label': 'Olahraga'},
    'https://www.cnnindonesia.com/nasional/20231108142334-617-1021530/
    ↳bobby-nasution-hadir-di-acara-deklarasi-dukung-prabowo-gibran': {'Media': '
    ↳CNN Indonesia', 'Label': 'Politik'},
    'https://www.cnnindonesia.com/nasional/20231108092234-617-1021356/
    ↳tkn-prabowo-sambut-baik-putusan-mkmk': {'Media': 'CNN Indonesia', 'Label': '
    ↳Politik'},
    'https://www.cnnindonesia.com/hiburan/20231108073048-220-1021340/
    ↳sinopsis-the-marvels-trio-superhero-perempuan-bersatu-lawan-dar-benn': '
    ↳{'Media': 'CNN Indonesia', 'Label': 'Hiburan'},
    'https://www.cnnindonesia.com/hiburan/20231108082925-220-1021350/
    ↳friends-alami-lonjakan-streaming-usai-kematian-matthew-perry': {'Media': '
    ↳CNN Indonesia', 'Label': 'Hiburan'},
    'https://www.liputan6.com/bola/read/5446924/
    ↳nostalgia-fiba-world-cup-2023-penjualan-final-dbl-jakarta-di-indonesia-arena-nyaris-ludes':
    ↳ {'Media': 'Liputan6', 'Label': 'Olahraga'},
    'https://www.liputan6.com/bola/read/5446920/
    ↳indonesia-bidik-tiket-olimpiade-paris-2024-di-kejuaraan-panjat-tebing-ifsc-asia-qualifier':
    ↳ {'Media': 'Liputan6', 'Label': 'Olahraga'},
    'https://www.liputan6.com/showbiz/read/5446901/
    ↳lee-junho-beri-bocoran-soal-fan-meeting-di-indonesia-pada-25-november': '
    ↳{'Media': 'Liputan6', 'Label': 'Hiburan'},
    'https://www.liputan6.com/showbiz/read/5444951/
    ↳ayu-ting-ting-pamer-momen-aktor-the-worst-of-evil-wi-ha-joon-terciduk-dua-kali-like-instagr
    ↳{'Media': 'Liputan6', 'Label': 'Hiburan'},
    'https://www.liputan6.com/news/read/5445021/
    ↳golkar-sebut-jokowi-prabowo-dan-gibran-akan-hadiri-acara-hut-ke-59-besok': '
    ↳{'Media': 'Liputan6', 'Label': 'Politik'},
    'https://www.liputan6.com/pemilu/read/5444959/
    ↳hasto-minta-tkd-ganjar-mahfud-gencarkan-sosialisasi-dansa-politik-nasional-jangan-jadi-peng
    ↳ {'Media': 'Liputan6', 'Label': 'Politik'},
    'https://entertainment.kompas.com/read/2023/11/07/155402166/
    ↳wyatt-russell-monarch-legacy-of-monsters-adalah-puncak-hiburan': {'Media': '
    ↳Kompas', 'Label': 'Hiburan'},

```

```

    'https://www.kompas.com/hype/read/2023/10/03/110546666/
    ↪jessica-mila-kurangi-pekerjaan-di-dunia-hiburan-dan-siapkan-diri-jadi-ibu':
    ↪{'Media': 'Kompas', 'Label': 'Hiburan'},
    'https://www.kompas.com/badminton/read/2023/11/08/07390068/
    ↪jadwal-korea-masters-2023--mulai-08.00-wib-tersaji-duel-merah-putih':␣
    ↪{'Media': 'Kompas', 'Label': 'Olahraga'},
    'https://www.kompas.com/badminton/read/2023/11/07/17132868/
    ↪kevin-rahmat-debut-di-korea-masters-2023-sempat-kagok-gim-kedua-lebih':␣
    ↪{'Media': 'Kompas', 'Label': 'Olahraga'},
    'https://nasional.kompas.com/read/2023/11/02/15415601/
    ↪jokowi-banyak-yang-berpikir-tahun-depan-ikn-jadi-siapa-yang-ngomong':␣
    ↪{'Media': 'Kompas', 'Label': 'Politik'},
    'https://nasional.kompas.com/read/2023/11/01/15434061/
    ↪jajaran-menteri-jokowi-yang-blak-blakan-dukung-prabowo-gibran-siapa-saja?
    ↪page=all': {'Media': 'Kompas', 'Label': 'Politik'},
    'https://www.cnbcindonesia.com/news/20210310131300-4-229236/
    ↪karaoke-bakal-buka-lagi-di-dki-pengusaha-habis-lebaran': {'Media': 'CNBC␣
    ↪Indonesia', 'Label': 'Hiburan'},
    'https://www.cnbcindonesia.com/tech/20231027061220-37-484093/
    ↪wota-bersiap-zee-freya-jkt48-bakal-ramaikan-shopee-live': {'Media': 'CNBC␣
    ↪Indonesia', 'Label': 'Hiburan'},
    'https://www.cnbcindonesia.com/news/20231023151723-4-482905/
    ↪wtc-mangga-dua-jadi-lapangan-bulutangkis-inikah-pemicunya': {'Media': 'CNBC␣
    ↪Indonesia', 'Label': 'Olahraga'},
    'https://www.cnbcindonesia.com/lifestyle/20231006090620-33-478386/
    ↪update-asian-games-2023-badminton-ri-cetak-sejarah-terburuk': {'Media':␣
    ↪'CNBC Indonesia', 'Label': 'Olahraga'},
    'https://www.cnbcindonesia.com/news/20231103135005-4-486172/
    ↪media-asing-soroti-politik-dinasti-di-asean-sebut-pemilu-ri': {'Media':␣
    ↪'CNBC Indonesia', 'Label': 'Politik'},
    'https://www.cnbcindonesia.com/news/20230810192036-4-462017/
    ↪punya-32-juta-pengikut-sosok-cawapres-anies-ini-buka-suara': {'Media': 'CNBC␣
    ↪Indonesia', 'Label': 'Politik'}
}

```

```

[4]: array = []

# Function mengambil text dari URL
def extract_text_content(url, class_name):
    response = requests.get(url)
    html_content = response.text
    soup = BeautifulSoup(html_content, 'html.parser')
    #kita harus mengambil teks yang didalam 'div' agar tidak mengambil teks␣
    ↪lainnya
    div = soup.find('div', class_=class_name)

```

```

if div:
    #jika textnya bukan berada di 'p' artinya kita tidak butuh text tersebut.
    paragraphs = div.find_all('p')
    paragraph_texts = [p.get_text(strip=True) for p in paragraphs]

    for p in paragraphs:
        paragraph_text = p.get_text(strip=True) + ' '
        paragraph_texts.append(paragraph_text)

    text_data = ' '.join(paragraph_texts)
else:
    text_data = ''

return text_data

for url, info in link.items():
    class_name = class_media(info['Media'])
    text_content = extract_text_content(url, class_name)
    #kita append agar bisa masukan ke csv
    array.append({'URL': url, 'Text': text_content, 'Media': info['Media'],
    ↪ 'Label': info['Label']})

```

```
[5]: data = pd.DataFrame(array)
```

```
[6]: data.head()
```

```
[6]:
```

	URL \	Text	Media	Label
0	https://www.cnnindonesia.com/olahraga/20231031...	Jonatan Christiemendapat banyak kritikan di 20...	CNN Indonesia	Olahraga
1	https://www.cnnindonesia.com/olahraga/20231105...	KemenanganDaejeon Jung KwanJang Red Sparksatas...	CNN Indonesia	Olahraga
2	https://www.cnnindonesia.com/nasional/20231108...	Wali Kota MedanBobby Nasutionhadir dalam acara...	CNN Indonesia	Politik
3	https://www.cnnindonesia.com/nasional/20231108...	Wakil Komandan Echo (Hukum dan Advokasi) Tim K...	CNN Indonesia	Politik
4	https://www.cnnindonesia.com/hiburan/202311080...	The Marvelsmerupakan film yang melanjutkan sal...	CNN Indonesia	Hiburan

```
[7]: data['Text'][12]
```

```
[7]: ' KOMPAS.com-Monarch: Legacy of Monstersakan menjadi serialspin-offdari waralaba
Godzilla yang bakal dirilis oleh Apple TV+ pada 17 November 2023. Serial
tersebut juga baru saja dirilis Inside Look melalui video yang diunggah oleh
akun YouTube Apple TV+ pada Minggu (5/11/2023) lalu. Pemeran karakter Lee Shaw
muda, Wyatt Russel, mengatakan Monarch: Legacy of Monsters merupakan sebuah
```

puncak hiburan. Baca juga: Monarch: Legacy of Monsters Tampilkan Kaiju-kaiju Baru "Ini adalah puncak hiburan, ada aksi dan kegembiraan," ujar Wyatt Russel dalam wawancaranya itu. Selain mengungkapkan hal tersebut, Wyatt juga menyatakan serial teranyarnya ini akan diisi oleh berbagai varian Kaiju baru. Sejauh yang diketahui, Kaiju yang menyerupai laba-laba raksasa hingga goliath bernapas es yang sebelumnya tak pernah muncul dalam waralaba Godzilla juga akan terlihat dalam serial ini. "Semua monster ini sangat degil, saya akan mengatakan itu," sambung aktor itu. Baca juga: Apple TV+ Keluarkan Trailer Resmi Monarch: Legacy of Monsters Selain Wyatt, wawancara tersebut juga menghadirkan Chris Black selaku produser eksekutif dari Monarch: Legacy of Monsters. Dalam wawancaranya, Black mengungkapkan serial Monarch garapannya itu akan membuat para penggemar baru tertarik untuk menyaksikannya. "Penggemar baru akan menonton ini dan berkata - wow, inilah cerita yang membuat saya tertarik," ujar produser tersebut. Baca juga: Jadwal Tayang, Pemeran, dan Alur Monarch: Legacy of Monsters Tak hanya bagi penggemar baru, Black juga mengatakan serialnya ini dapat membuat para pengikut Monsterverse dapat mengenali sebuah dunia yang akrab dengan mereka. "Penonton yang menyukai Monsterverse akan melihat dunia yang mereka kenali" ungkap Chris Black. "Merupakan sebuah kehormatan untuk diundang oleh teman dan kolaborator kami di Toho dan Legendary untuk bermain di alam Monsterverse ini" tutupnya. KOMPAS.com-Monarch: Legacy of Monsters akan menjadi serial spin-off dari waralaba Godzilla yang bakal dirilis oleh Apple TV+ pada 17 November 2023. Serial tersebut juga baru saja merilis Inside Look melalui video yang diunggah oleh akun YouTube Apple TV+ pada Minggu (5/11/2023) lalu. Pemeran karakter Lee Shaw muda, Wyatt Russel, mengatakan Monarch: Legacy of Monsters merupakan sebuah puncak hiburan. Baca juga: Monarch: Legacy of Monsters Tampilkan Kaiju-kaiju Baru "Ini adalah puncak hiburan, ada aksi dan kegembiraan," ujar Wyatt Russel dalam wawancaranya itu. Selain mengungkapkan hal tersebut, Wyatt juga menyatakan serial teranyarnya ini akan diisi oleh berbagai varian Kaiju baru. Sejauh yang diketahui, Kaiju yang menyerupai laba-laba raksasa hingga goliath bernapas es yang sebelumnya tak pernah muncul dalam waralaba Godzilla juga akan terlihat dalam serial ini. "Semua monster ini sangat degil, saya akan mengatakan itu," sambung aktor itu. Baca juga: Apple TV+ Keluarkan Trailer Resmi Monarch: Legacy of Monsters Selain Wyatt, wawancara tersebut juga menghadirkan Chris Black selaku produser eksekutif dari Monarch: Legacy of Monsters. Dalam wawancaranya, Black mengungkapkan serial Monarch garapannya itu akan membuat para penggemar baru tertarik untuk menyaksikannya. "Penggemar baru akan menonton ini dan berkata - wow, inilah cerita yang membuat saya tertarik," ujar produser tersebut. Baca juga: Jadwal Tayang, Pemeran, dan Alur Monarch: Legacy of Monsters Tak hanya bagi penggemar baru, Black juga mengatakan serialnya ini dapat membuat para pengikut Monsterverse dapat mengenali sebuah dunia yang akrab dengan mereka. "Penonton yang menyukai Monsterverse akan melihat dunia yang mereka kenali" ungkap Chris Black. "Merupakan sebuah kehormatan untuk diundang oleh teman dan kolaborator kami di Toho dan Legendary untuk bermain di alam Monsterverse ini" tutupnya. '

#Data Preprocessing

```
[8]: def cleansing(text):

    text = text.lower()

    text = re.compile('%s' % re.escape(string.punctuation)).sub(' ', text)
    text = re.sub(r'\[[0-9]*\]', ' ', text)
    text = re.sub(r'~\w\s', ' ', text)
    text = re.sub(r'\s+', ' ', text)
    text = re.sub(r'\d', ' ', text)

    words = text.split()
    text = ' '.join(words)

    text = text.replace('\n', '')
    text = text.replace('advertisement', '')
    text = text.replace('scroll to resume content', '')
    text = text.replace('scroll to continue with content', '')
    text = text.replace('liputan com jakarta', '')
    text = text.replace('kompas com ', '')
    text = text.replace('jakarta cnbc indonesia ', '')

    return text
```

```
[9]: nltk.download('stopwords')
nltk.download('wordnet')
nltk.download('punkt')
```

```
[nltk_data] Downloading package stopwords to /root/nltk_data...
[nltk_data] Unzipping corpora/stopwords.zip.
[nltk_data] Downloading package wordnet to /root/nltk_data...
[nltk_data] Downloading package punkt to /root/nltk_data...
[nltk_data] Unzipping tokenizers/punkt.zip.
```

```
[9]: True
```

```
[10]: data['Text'] = data['Text'].apply(cleansing)
```

```
[11]: data['Text'][12]
```

```
[11]: 'monarch legacy of monsters akan menjadi serial spin off dari waralaba godzilla
yang bakal dirilis oleh apple tv pada november serial tersebut juga baru saja
merilis inside look melalui video yang diunggah oleh akun youtube apple tv pada
minggu lalu pemeran karakter lee shaw muda wyatt russel mengatakan monarch
legacy of monsters merupakan sebuah puncak hiburan baca juga monarch legacy of
monsters tampilkan kaiju kaiju baru ini adalah puncak hiburan ada aksi dan
kegembiraan ujar wyatt russel dalam wawancaranya itu selain mengungkapkan hal
tersebut wyatt juga menyatakan serial teranyarnya ini akan diisi oleh berbagai
```

varian kaiju baru sejauh yang diketahui kaiju yang menyerupai laba laba raksasa hingga goliath bernapas es yang sebelumnya tak pernah muncul dalam waralaba godzilla juga akan terlihat dalam serial ini semua monster ini sangat degil saya akan mengatakan itu sambung aktor itu baca juga apple tv keluarkan trailer resmi monarch legacy of monsters selain wyatt wawancara tersebut juga menghadirkan chris black selaku produser eksekutif dari monarch legacy of monsters dalam wawancaranya black mengungkapkan serial monarch garapannya itu akan membuat para penggemar baru tertarik untuk menyaksikannya penggemar baru akan menonton ini dan berkata wow inilah cerita yang membuat saya tertarik ujar produser tersebut baca juga jadwal tayang pemeran dan alur monarch legacy of monsters tak hanya bagi penggemar baru black juga mengatakan serialnya ini dapat membuat para pengikutmonsterversedapat mengenali sebuah dunia yang akrab dengan mereka penonton yang menyukai monsterverse akan melihat dunia yang mereka kenali ungkap chris black merupakan sebuah kehormatan untuk diundang oleh teman dan kolaborator kami di toho dan legendary untuk bermain di alam monsterverse ini tutupnya monarch legacy of monsters akan menjadi serialspin off dari waralaba godzilla yang bakal dirilis oleh apple tv pada november serial tersebut juga baru saja dirilis inside look melalui video yang diunggah oleh akun youtube apple tv pada minggu lalu pemeran karakter lee shaw muda wyatt russel mengatakan monarch legacy of monsters merupakan sebuah puncak hiburan baca juga monarch legacy of monsters tampilkan kaiju kaiju baru ini adalah puncak hiburan ada aksi dan kegembiraan ujar wyatt russel dalam wawancaranya itu selain mengungkapkan hal tersebut wyatt juga menyatakan serial teranyarnya ini akan diisi oleh berbagai varian kaiju baru sejauh yang diketahui kaiju yang menyerupai laba laba raksasa hingga goliath bernapas es yang sebelumnya tak pernah muncul dalam waralaba godzilla juga akan terlihat dalam serial ini semua monster ini sangat degil saya akan mengatakan itu sambung aktor itu baca juga apple tv keluarkan trailer resmi monarch legacy of monsters selain wyatt wawancara tersebut juga menghadirkan chris black selaku produser eksekutif dari monarch legacy of monsters dalam wawancaranya black mengungkapkan serial monarch garapannya itu akan membuat para penggemar baru tertarik untuk menyaksikannya penggemar baru akan menonton ini dan berkata wow inilah cerita yang membuat saya tertarik ujar produser tersebut baca juga jadwal tayang pemeran dan alur monarch legacy of monsters tak hanya bagi penggemar baru black juga mengatakan serialnya ini dapat membuat para pengikutmonsterversedapat mengenali sebuah dunia yang akrab dengan mereka penonton yang menyukai monsterverse akan melihat dunia yang mereka kenali ungkap chris black merupakan sebuah kehormatan untuk diundang oleh teman dan kolaborator kami di toho dan legendary untuk bermain di alam monsterverse ini tutupnya'

```
[12]: data.to_csv('berita.csv', index=False)
      df = pd.read_csv('berita.csv')
```

```
[13]: print(df)
```

URL \

0 <https://www.cnnindonesia.com/olahraga/20231031...>

1 <https://www.cnnindonesia.com/olahraga/20231105...>  
2 <https://www.cnnindonesia.com/nasional/20231108...>  
3 <https://www.cnnindonesia.com/nasional/20231108...>  
4 <https://www.cnnindonesia.com/hiburan/202311080...>  
5 <https://www.cnnindonesia.com/hiburan/202311080...>  
6 <https://www.liputan6.com/bola/read/5446924/nos...>  
7 <https://www.liputan6.com/bola/read/5446920/ind...>  
8 <https://www.liputan6.com/showbiz/read/5446901/...>  
9 <https://www.liputan6.com/showbiz/read/5444951/...>  
10 <https://www.liputan6.com/news/read/5445021/gol...>  
11 <https://www.liputan6.com/pemilu/read/5444959/h...>  
12 <https://entertainment.kompas.com/read/2023/11/...>  
13 <https://www.kompas.com/hype/read/2023/10/03/11...>  
14 <https://www.kompas.com/badminton/read/2023/11/...>  
15 <https://www.kompas.com/badminton/read/2023/11/...>  
16 <https://nasional.kompas.com/read/2023/11/02/15...>  
17 <https://nasional.kompas.com/read/2023/11/01/15...>  
18 <https://www.cnbcindonesia.com/news/20210310131...>  
19 <https://www.cnbcindonesia.com/tech/20231027061...>  
20 <https://www.cnbcindonesia.com/news/20231023151...>  
21 <https://www.cnbcindonesia.com/lifestyle/202310...>  
22 <https://www.cnbcindonesia.com/news/20231103135...>  
23 <https://www.cnbcindonesia.com/news/20230810192...>

	Text	Media \
0	jonatan christiemendapat banyak kritikan di pa...	CNN Indonesia
1	kemenangandaejeon jung kwanjang red sparksatas...	CNN Indonesia
2	wali kota medanbobby nasutionhadir dalam acara...	CNN Indonesia
3	wakil komandan echo hukum dan advokasi tim kam...	CNN Indonesia
4	the marvelsmerupakan film yang melanjutkan sal...	CNN Indonesia
5	sitkomfriendsmengalami pelonjakan penonton di ...	CNN Indonesia
6	pecinta bola basket di jakarta dan sekitarnya...	Liputan6
7	federasi panjat tebing indonesia fpti bakal m...	Liputan6
8	kurang dari sebulanlee junhoakan ke indonesia ...	Liputan6
9	sudah banyak interaksi unik antara artis indo...	Liputan6
10	partaigolkarakaran memperingati hari ulang tahun...	Liputan6
11	sekjenpdiphasto kristiyanto meminta tim peme...	Liputan6
12	monarch legacy of monstersakan menjadi serials...	Kompas
13	jakarta aktrisjessica milamengaku sempat menga...	Kompas
14	jadwalkorea masters hari ini rabu akan berlang...	Kompas
15	ganda putra indonesia kevin sanjaya sukamuljo ...	Kompas
16	jakarta presiden joko widodomembantah anggapan...	Kompas
17	jakarta bakal calon presiden capres dan calon ...	Kompas
18	pemerintah provinsi dki jakarta di bawah guber...	CNBC Indonesia
19	kabar gembira bagi penggemar jkt atau wota kar...	CNBC Indonesia
20	bukan menambah arena tenant pusat perbelanjaan...	CNBC Indonesia
21	kontingen indonesia tercatat masih belum berha...	CNBC Indonesia
22	isu politik dinasti di asia tenggara menjadi s...	CNBC Indonesia



23 salah satu kandidat bakal calon wakil presiden... CNBC Indonesia

	Label
0	Olahraga
1	Olahraga
2	Politik
3	Politik
4	Hiburan
5	Hiburan
6	Olahraga
7	Olahraga
8	Hiburan
9	Hiburan
10	Politik
11	Politik
12	Hiburan
13	Hiburan
14	Olahraga
15	Olahraga
16	Politik
17	Politik
18	Hiburan
19	Hiburan
20	Olahraga
21	Olahraga
22	Politik
23	Politik

#Stopwords and Lemmatization

```
[14]: from nltk.tokenize import word_tokenize
      from nltk.corpus import stopwords

      stop_words = set(stopwords.words('indonesian'))
      lemmatizer = WordNetLemmatizer()

      def lemmatize_and_stopwords(text):
          tokens = nltk.word_tokenize(text)
          lemmatized_tokens = [lemmatizer.lemmatize(word) for word in tokens if word
                               ↪not in stop_words]
          lemmatized_text = ' '.join(lemmatized_tokens)
          return lemmatized_text

      df['Tokenized Text'] = df['Text'].apply(lambda text: [lemmatizer.
          ↪lemmatize(word) for word in nltk.word_tokenize(text) if word not in
          ↪stop_words])
      df['Lemmatized Text'] = df['Text'].apply(lemmatize_and_stopwords)
```

```
[15]: df.head()
```

```
[15]: URL \
0 https://www.cnnindonesia.com/olahraga/20231031...
1 https://www.cnnindonesia.com/olahraga/20231105...
2 https://www.cnnindonesia.com/nasional/20231108...
3 https://www.cnnindonesia.com/nasional/20231108...
4 https://www.cnnindonesia.com/hiburan/202311080...

Text Media Label \
0 jonatan christiemendapat banyak kritikan di pa... CNN Indonesia Olahraga
1 kemenangandaejeon jung kwanjang red sparksatas... CNN Indonesia Olahraga
2 wali kota medanbobby nasutionhadir dalam acara... CNN Indonesia Politik
3 wakil komandan echo hukum dan advokasi tim kam... CNN Indonesia Politik
4 the marvelsmerupakan film yang melanjutkan sal... CNN Indonesia Hiburan

Tokenized Text \
0 [jonatan, christiemendapat, kritikan, merujuk,...
1 [kemenangandaejeon, jung, kwanjang, red, spark...
2 [wali, kota, medanbobby, nasutionhadir, acara,...
3 [wakil, komandan, echo, hukum, advokasi, tim, ...
4 [the, marvelsmerupakan, film, melanjutkan, sal...

Lemmatized Text
0 jonatan christiemendapat kritikan merujuk stat...
1 kemenangandaejeon jung kwanjang red sparksatas...
2 wali kota medanbobby nasutionhadir acara dekla...
3 wakil komandan echo hukum advokasi tim kampanye...
4 the marvelsmerupakan film melanjutkan salah ki...
```

```
[16]: df['Tokenized Text']
```

```
[16]: 0 [jonatan, christiemendapat, kritikan, merujuk,...
1 [kemenangandaejeon, jung, kwanjang, red, spark...
2 [wali, kota, medanbobby, nasutionhadir, acara,...
3 [wakil, komandan, echo, hukum, advokasi, tim, ...
4 [the, marvelsmerupakan, film, melanjutkan, sal...
5 [sitkomfriendsmengalami, pelonjakan, penonton,...
6 [pecinta, bola, basket, jakarta, antusias, men...
7 [federasi, panjat, tebing, indonesia, fpti, me...
8 [sebulanlee, junhoakan, indonesia, menggelar, ...
9 [interaksi, unik, artis, indonesia, artis, neg...
10 [partaigolkarakaran, memperingati, ulang, hut, s...
11 [sekjenpdiphasto, kristiyanto, tim, kemenangan...
12 [monarch, legacy, of, monstersakan, serialspin...
13 [jakarta, aktrisjessica, milamengaku, mengalam...
14 [jadwalkorea, master, rabu, gwangju, woman, s,...
```

```

15 [ganda, putra, indonesia, kevin, sanjaya, suka...
16 [jakarta, presiden, joko, widodomembantah, ang...
17 [jakarta, calon, presiden, capres, calon, waki...
18 [pemerintah, provinsi, dki, jakarta, gubernur,...
19 [kabar, gembira, penggemar, jkt, wota, idol, z...
20 [menambah, arena, tenant, pusat, perbelanjaan,...
21 [kontingen, indonesia, tercatat, berhasil, men...
22 [isu, politik, dinasti, asia, tenggara, sorota...
23 [salah, kandidat, calon, wakil, presiden, baca...
Name: Tokenized Text, dtype: object

```

```
[17]: df['Lemmatized Text']
```

```

[17]: 0 jonatan christiemendapat kritikan merujuk stat...
1 kemenangandaejeon jung kwanjang red sparksatas...
2 wali kota medanbobby nasutionhadir acara dekla...
3 wakil komandan echo hukum advokasi tim kampany...
4 the marvelsmerupakan film melanjutkan salah ki...
5 sitkomfriendsmengalami pelonjakan penonton lay...
6 pecinta bola basket jakarta antusias menyambut...
7 federasi panjat tebing indonesia fpti menggela...
8 sebulanlee junhoakan indonesia menggelar fan m...
9 interaksi unik artis indonesia artis negeri gi...
10 partaigolkarakan memperingati ulang hut senin ...
11 sekjenpdiphasto kristiyanto tim kemenangan dae...
12 monarch legacy of monstersakan serialspin offd...
13 jakarta aktrisjessica milamengaku mengalamimor...
14 jadwalkorea master rabu gwangju woman s univer...
15 ganda putra indonesia kevin sanjaya sukamuljo ...
16 jakarta presiden joko widodomembantah anggapan...
17 jakarta calon presiden capres calon wakil pres...
18 pemerintah provinsi dki jakarta gubernur anies...
19 kabar gembira penggemar jkt wota idol zee frey...
20 menambah arena tenant pusat perbelanjaan wtc m...
21 kontingen indonesia tercatat berhasil menambah...
22 isu politik dinasti asia tenggara sorotan bera...
23 salah kandidat calon wakil presiden bacawapres...
Name: Lemmatized Text, dtype: object

```

```
[18]: df.info()
```

```

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 24 entries, 0 to 23
Data columns (total 6 columns):
#   Column                Non-Null Count  Dtype
---  -
0   URL                    24 non-null    object

```

```

1  Text          24 non-null  object
2  Media         24 non-null  object
3  Label         24 non-null  object
4  Tokenized Text 24 non-null  object
5  Lemmatized Text 24 non-null  object

```

dtypes: object(6)

memory usage: 1.2+ KB

#Modelling

```

[19]: from sklearn.feature_extraction.text import CountVectorizer
from sklearn.feature_extraction.text import TfidfVectorizer
X_train, X_test, y_train, y_test = train_test_split(df['Lemmatized Text'],
↳df['Label'], test_size=0.2, random_state=42)

# Metode 1: TF-IDF
tfidf_vectorizer = TfidfVectorizer(max_features=50, min_df=3)
X_train_tfidf = tfidf_vectorizer.fit_transform(X_train)
X_test_tfidf = tfidf_vectorizer.transform(X_test)

# Melatih model SVM dengan metode TF-IDF
svm_classifier_tfidf = SVC(kernel='linear')
svm_classifier_tfidf.fit(X_train_tfidf, y_train)
svm_predictions_tfidf = svm_classifier_tfidf.predict(X_test_tfidf)

# Melatih model Random Forest dengan metode TF-IDF
rf_classifier_tfidf = RandomForestClassifier(n_estimators=100, random_state=42)
rf_classifier_tfidf.fit(X_train_tfidf, y_train)
rf_predictions_tfidf = rf_classifier_tfidf.predict(X_test_tfidf)

# Evaluation
print("Hasil SVM dengan TF-IDF:")
print(accuracy_score(y_test, svm_predictions_tfidf))
print(classification_report(y_test, svm_predictions_tfidf))

print("Hasil Random Forest dengan TF-IDF:")
print(accuracy_score(y_test, rf_predictions_tfidf))
print(classification_report(y_test, rf_predictions_tfidf))

```

Hasil SVM dengan TF-IDF:

0.8

	precision	recall	f1-score	support
Hiburan	1.00	0.50	0.67	2
Olahraga	0.50	1.00	0.67	1
Politik	1.00	1.00	1.00	2
accuracy			0.80	5

macro avg	0.83	0.83	0.78	5
weighted avg	0.90	0.80	0.80	5

Hasil Random Forest dengan TF-IDF:  
0.6

	precision	recall	f1-score	support
Hiburan	0.50	0.50	0.50	2
Olahraga	0.50	1.00	0.67	1
Politik	1.00	0.50	0.67	2
accuracy			0.60	5
macro avg	0.67	0.67	0.61	5
weighted avg	0.70	0.60	0.60	5

Reference TF-IDF: <https://www.kaggle.com/code/sainijagjit/text-classification-using-svm>

```
[20]: # Metode 2: Count Vectorizer
count_vectorizer = CountVectorizer(max_features=50, min_df=3)
X_train_count = count_vectorizer.fit_transform(X_train)
X_test_count = count_vectorizer.transform(X_test)

# Melatih model SVM dengan metode Count Vectorizer
svm_classifier_count = SVC(kernel='linear')
svm_classifier_count.fit(X_train_count, y_train)
svm_predictions_count = svm_classifier_count.predict(X_test_count)

# Melatih model Random Forest dengan metode Count Vectorizer
rf_classifier_count = RandomForestClassifier(n_estimators=100, random_state=42)
rf_classifier_count.fit(X_train_count, y_train)
rf_predictions_count = rf_classifier_count.predict(X_test_count)

#Evaluation
print("Hasil SVM dengan Count Vectorizer:")
print(accuracy_score(y_test, svm_predictions_count))
print(classification_report(y_test, svm_predictions_count))

print("Hasil Random Forest dengan Count Vectorizer:")
print(accuracy_score(y_test, rf_predictions_count))
print(classification_report(y_test, rf_predictions_count))
```

Hasil SVM dengan Count Vectorizer:  
0.8

	precision	recall	f1-score	support
Hiburan	0.67	1.00	0.80	2
Olahraga	0.00	0.00	0.00	1

Politik	1.00	1.00	1.00	2
accuracy			0.80	5
macro avg	0.56	0.67	0.60	5
weighted avg	0.67	0.80	0.72	5

Hasil Random Forest dengan Count Vectorizer:  
0.8

	precision	recall	f1-score	support
Hiburan	1.00	0.50	0.67	2
Olahraga	0.50	1.00	0.67	1
Politik	1.00	1.00	1.00	2
accuracy			0.80	5
macro avg	0.83	0.83	0.78	5
weighted avg	0.90	0.80	0.80	5

```
/usr/local/lib/python3.10/dist-packages/sklearn/metrics/_classification.py:1344:
UndefinedMetricWarning: Precision and F-score are ill-defined and being set to
0.0 in labels with no predicted samples. Use `zero_division` parameter to
control this behavior.
```

```
_warn_prf(average, modifier, msg_start, len(result))
```

```
/usr/local/lib/python3.10/dist-packages/sklearn/metrics/_classification.py:1344:
UndefinedMetricWarning: Precision and F-score are ill-defined and being set to
0.0 in labels with no predicted samples. Use `zero_division` parameter to
control this behavior.
```

```
_warn_prf(average, modifier, msg_start, len(result))
```

```
/usr/local/lib/python3.10/dist-packages/sklearn/metrics/_classification.py:1344:
UndefinedMetricWarning: Precision and F-score are ill-defined and being set to
0.0 in labels with no predicted samples. Use `zero_division` parameter to
control this behavior.
```

```
_warn_prf(average, modifier, msg_start, len(result))
```

#Predicting Text

```
[21]: predict = 'Ketua DPP Partai Golkar Ace Hasan Syadzily membalas tanggapan_
↳Sekretaris Tim koordinasi Relawan Pemenangan Pilpres 2024 PDI-P(TKRPP PDI-P)_
↳Deddy Yevri Sitorus yang menuding drama politik seperti sinetron justru_
↳banyak terjadi di lingkungan Presiden Joko Widodo (Jokowi). Ace meminta_
↳Deddy tidak membawa perasaan atas ucapan Jokowi yang menyinggung soal drama_
↳politik beberapa waktu terakhir.'
#https://nasional.kompas.com/read/2023/11/07/13455011/
↳golkar-bela-jokowi-yang-disindir-pdi-p-enggak-usah-baper
predict = cleansing(predict)

preprocessed_predict = lemmatize_and_stopwords(predict)
```

```
[22]: preprocessed_predict
```

```
[22]: 'ketua dpp partai golkar ace hasan syadzily membalas tanggapan sekretaris tim
koordinasi relawan pemenangan pilpres pdi p tkrpp pdi p deddy yevri sitorus
menuding drama politik sinetron lingkungan presiden joko widodo jokowi ace deddy
membawa perasaan ucapan jokowi menyinggung drama politik'
```

```
[23]: predict_tfidf_vectorized = tfidf_vectorizer.transform([preprocessed_predict])
      ↪# Use your pre-trained TF-IDF model

# TF-IDF
prediction_rf_tfidf = rf_classifier_tfidf.predict(predict_tfidf_vectorized)
      ↪Use your pre-trained classifier
prediction_svm_tfidf = svm_classifier_tfidf.predict(predict_tfidf_vectorized)
print('TF-IDF:')
print(f'Random Forests Model Predicted Category: {prediction_rf_tfidf}')
print(f'SVM Model Predicted Category: {prediction_svm_tfidf}')
print('')

predict_count_vectorized = count_vectorizer.transform([preprocessed_predict])
      ↪# Use your pre-trained TF-IDF model

# CV
prediction_rf_count = rf_classifier_count.predict(predict_count_vectorized)
      ↪Use your pre-trained classifier
prediction_svm_count = svm_classifier_count.predict(predict_count_vectorized)
print('Count Vectorizer:')
print(f'Random Forests Model Predicted Category: {prediction_rf_count}')
print(f'SVM Model Predicted Category: {prediction_svm_count}')
```

TF-IDF:

Random Forests Model Predicted Category: ['Politik']

SVM Model Predicted Category: ['Politik']

Count Vectorizer:

Random Forests Model Predicted Category: ['Olahraga']

SVM Model Predicted Category: ['Hiburan']

```
[24]: predict_link = 'https://www.cnnindonesia.com/hiburan/20231108154726-220-1021598/
      ↪5-perbedaan-manga-dan-anime-attack-on-titan-final-season-part-4'
predict_link_media = 'CNN Indonesia'

class_name2 = class_media(predict_link_media)
text_content2 = extract_text_content(predict_link, class_name2)
```

```
[25]: class_name2
```

[25]: 'detail-text'

[26]: text\_content2

[26]: 'Artikel ini mengandung spoiler/beberan... Attack on Titan benar-benar tamat usai animenya tamat pada 4 November lalu menyusul manganya yang tutup buku terlebih dulu pada 2021. Anime Attack on Titan dinilai sangat setia dalam mengikuti cerita versi manga. Sehingga, tidak ada perbedaan signifikan dalam alur ceritanya. ADVERTISEMENT SCROLL TO CONTINUE WITH CONTENT Meski demikian, episode pemungkas Attack on Titan The Final Season Part 4 ternyata memiliki sedikit perbedaan dari karya aslinya. Salah satunya adalah percakapan terakhir antara Eren dengan salah satu sahabatnya, Armin. Hajime Isayama selaku kreator Attack on Titan memperbaiki dialog kedua karakter tersebut setelah dikritik oleh para penggemar. Armin bertemu dengan Eren dalam sebuah memori setelah leher sahabatnya itu ditebas oleh Mikasa. Dalam memori itu, Eren menjelaskan kepada Armin alasannya melakukan Guncangan Bumi. Versi manga Attack on Titan, Armin kemudian kemudian berterima kasih kepada Eren karena rela menjadi penjahat demi menyelamatkan bangsa Eldia. "Eren, terima kasih," kata Armin dalam versi manga. "Kamu menjadi pembunuh massal demi kami. "Aku berjanji tidak akan membiarkan kesalahan besar yang kamu buat menjadi sia-sia," sambung Armin. Percakapan ini kemudian diperbaiki dan diperpanjang dalam versi anime. Eren menjelaskan rencananya untuk menginjak 80 persen umat manusia kepada Armin. Armin pun berusaha untuk menghentikan niat sahabatnya itu sebelum terlambat, tapi Eren mengatakan rencana itu sudah pasti akan terjadi. Armin tetap berterima kasih kepada Eren. Bukan karena sahabatnya menjadi pembunuh, tapi karena telah menunjukkan dunia luar di balik tembok pulau Paradis. "Terima kasih Eren karena telah menunjukkan apa yang ada di balik tembok, karena telah menunjukkan kepadaku pemandangan ini," kata Armin kepada Eren. Masih di momen Eren dan Armin mengobrol dalam memori. Obrolan sepasang sahabat itu terlihat selalu berubah latar dan waktu, dimulai dari ketika mereka masih bocah hingga remaja. Di momen terakhir, keduanya berbincang di tanah hasil injakan ribuan titan yang menewaskan manusia tanpa pandang bulu. Ini terlihat dalam versi manga. Namun, versi anime menunjukkan pemandangan yang sedikit berbeda, yaitu hasil injakan kaki titan itu digenangi darah manusia yang menjadi korban Eren. Eren kemudian mengambil segenggam rambut dan tulang manusia. Ia tampak menyesali perbuatannya. Sedangkan Armin mengambil kerang yang kemudian diberikan kepada Eren dan berjanji untuk menyusul Eren di neraka. Lanjut ke sebelah... Artikel ini mengandung spoiler/beberan... Attack on Titan benar-benar tamat usai animenya tamat pada 4 November lalu menyusul manganya yang tutup buku terlebih dulu pada 2021. Anime Attack on Titan dinilai sangat setia dalam mengikuti cerita versi manga. Sehingga, tidak ada perbedaan signifikan dalam alur ceritanya. ADVERTISEMENT SCROLL TO CONTINUE WITH CONTENT Meski demikian, episode pemungkas Attack on Titan The Final Season Part 4 ternyata memiliki sedikit perbedaan dari karya aslinya. Salah satunya adalah percakapan terakhir antara Eren dengan salah satu sahabatnya, Armin. Hajime Isayama selaku kreator Attack on Titan memperbaiki dialog kedua karakter tersebut setelah dikritik oleh para penggemar. Armin bertemu dengan Eren dalam sebuah memori setelah leher



sahabatnya itu ditebas oleh Mikasa. Dalam memori itu, Eren menjelaskan kepada Armin alasannya melakukan Guncangan Bumi. Versi manga Attack on Titan, Armin kemudian kemudian berterima kasih kepada Eren karena rela menjadi penjahat demi menyelamatkan bangsa Eldia. "Eren, terima kasih," kata Armin dalam versi manga. "Kamu menjadi pembunuh massal demi kami. "Aku berjanji tidak akan membiarkan kesalahan besar yang kamu buat menjadi sia-sia," sambung Armin. Percakapan ini kemudian diperbaiki dan diperpanjang dalam versi anime. Eren menjelaskan rencananya untuk menginjak 80 persen umat manusia kepada Armin. Armin pun berusaha untuk menghentikan niat sahabatnya itu sebelum terlambat, tapi Eren mengatakan rencana itu sudah pasti akan terjadi. Armin tetap berterima kasih kepada Eren. Bukan karena sahabatnya menjadi pembunuh, tapi karena telah menunjukkan dunia luar di balik tembok pulau Paradis. "Terima kasih Eren karena telah menunjukkan apa yang ada di balik tembok, karena telah menunjukkan kepadaku pemandangan ini," kata Armin kepada Eren. Masih di momen Eren dan Armin mengobrol dalam memori. Obrolan sepasang sahabat itu terlihat selalu berubah latar dan waktu, dimulai dari ketika mereka masih bocah hingga remaja. Di momen terakhir, keduanya berbincang di tanah hasil injakan ribuan titan yang menewaskan manusia tanpa pandang bulu. Ini terlihat dalam versi manga. Namun, versi anime menunjukkan pemandangan yang sedikit berbeda, yaitu hasil injakan kaki titan itu digenangi darah manusia yang menjadi korban Eren. Eren kemudian mengambil segenggam rambut dan tulang manusia. Ia tampak menyesali perbuatannya. Sedangkan Armin mengambil kerang yang kemudian diberikan kepada Eren dan berjanji untuk menyusul Eren di neraka. Lanjut ke sebelah... '

```
[27]: data_predict = cleansing(text_content2)
preprocessed_predict2 = lemmatize_and_stopwords(data_predict)
predict_tfidf_vectorized2 = tfidf_vectorizer.transform([preprocessed_predict2])

# TF-IDF
prediction_rf_tfidf2 = rf_classifier_tfidf.predict(predict_tfidf_vectorized2)
prediction_svm_tfidf2 = svm_classifier_tfidf.predict(predict_tfidf_vectorized2)
print('TF-IDF:')
print(f'Random Forests Model Predicted Category: {prediction_rf_tfidf2}')
print(f'SVM Model Predicted Category: {prediction_svm_tfidf2}')
print('')

predict_count_vectorized2 = count_vectorizer.transform([preprocessed_predict2])

# CV
prediction_rf_count2 = rf_classifier_count.predict(predict_count_vectorized2)
prediction_svm_count2 = svm_classifier_count.predict(predict_count_vectorized2)
print('Count Vectorizer:')
print(f'Random Forests Model Predicted Category: {prediction_rf_count2}')
print(f'SVM Model Predicted Category: {prediction_svm_count2}')
```

TF-IDF:

Random Forests Model Predicted Category: ['Hiburan']

SVM Model Predicted Category: ['Hiburan']

Count Vectorizer:

Random Forests Model Predicted Category: ['Hiburan']

SVM Model Predicted Category: ['Hiburan']

Reference: <https://www.kaggle.com/code/adamschroeder/countvectorizer-tfidfvectorizer-predict-comments>

[28]: data\_predict

[28]: 'artikel ini mengandung spoiler beberan attack on titan benar benar tamat usai animenya tamat pada november lalu menyusul manganya yang tutup buku terlebih dulu pada anime attack on titan dinilai sangat setia dalam mengikuti cerita versi manga sehingga tidak ada perbedaan signifikan dalam alur ceritanya meski demikian episode pemungkas attack on titan the final season part ternyata memiliki sedikit perbedaan dari karya aslinya salah satunya adalah percakapan terakhir antara eren dengan salah satu sahabatnya armin hajime isayama selaku kreator attack on titan memperbaiki dialog kedua karakter tersebut setelah dikritik oleh para penggemar armin bertemu dengan eren dalam sebuah memori setelah leher sahabatnya itu ditebas oleh mikasa dalam memori itu eren menjelaskan kepada armin alasannya melakukan guncangan bumi versi manga attack on titan armin kemudian kemudian berterima kasih kepada eren karena rela menjadi penjahat demi menyelamatkan bangsa eldia eren terima kasih kata armin dalam versi manga kamu menjadi pembunuh massal demi kami aku berjanji tidak akan membiarkan kesalahan besar yang kamu buat menjadi sia sia sambung armin percakapan ini kemudian diperbaiki dan diperpanjang dalam versi anime eren menjelaskan rencananya untuk menginjak persen umat manusia kepada armin armin pun berusaha untuk menghentikan niat sahabatnya itu sebelum terlambat tapi eren mengatakan rencana itu sudah pasti akan terjadi armin tetap berterima kasih kepada eren bukan karena sahabatnya menjadi pembunuh tapi karena telah menunjukkan dunia luar di balik tembok pulau paradisi terima kasih eren karena telah menunjukkan apa yang ada di balik tembok karena telah menunjukkan kepadaku pemandangan ini kata armin kepada eren masih di momen eren dan armin mengobrol dalam memori obrolan sepasang sahabat itu terlihat selalu berubah latar dan waktu dimulai dari ketika mereka masih bocah hingga remaja di momen terakhir keduanya berbincang di tanah hasil injakan ribuan titan yang menewaskan manusia tanpa pandang bulu ini terlihat dalam versi manga namun versi anime menunjukkan pemandangan yang sedikit berbeda yaitu hasil injakan kaki titan itu digenangi darah manusia yang menjadi korban eren eren kemudian mengambil segenggam rambut dan tulang manusia ia tampak menyesali perbuatannya sedangkan armin mengambil kerang yang kemudian diberikan kepada eren dan berjanji untuk menyusul eren di neraka lanjut ke sebelah artikel ini mengandung spoiler beberan attack on titan benar benar tamat usai animenya tamat pada november lalu menyusul manganya yang tutup buku terlebih dulu pada anime attack on titan dinilai sangat setia dalam mengikuti cerita versi manga sehingga tidak ada perbedaan signifikan dalam alur ceritanya meski demikian episode pemungkas attack on titan the final season part ternyata memiliki sedikit perbedaan dari karya aslinya salah satunya

adalah percakapan terakhir antara eren dengan salah satu sahabatnya armin hajime isayama selaku kreator attack on titan memperbaiki dialog kedua karakter tersebut setelah dikritik oleh para penggemar armin bertemu dengan eren dalam sebuah memori setelah leher sahabatnya itu ditebas oleh mikasa dalam memori itu eren menjelaskan kepada armin alasannya melakukan guncangan bumi versi manga attack on titan armin kemudian kemudian berterima kasih kepada eren karena rela menjadi penjahat demi menyelamatkan bangsa eldia eren terima kasih kata armin dalam versi manga kamu menjadi pembunuh massal demi kami aku berjanji tidak akan membiarkan kesalahan besar yang kamu buat menjadi sia sia sambung armin percakapan ini kemudian diperbaiki dan diperpanjang dalam versi anime eren menjelaskan rencananya untuk menginjak persen umat manusia kepada armin armin pun berusaha untuk menghentikan niat sahabatnya itu sebelum terlambat tapi eren mengatakan rencana itu sudah pasti akan terjadi armin tetap berterima kasih kepada eren bukan karena sahabatnya menjadi pembunuh tapi karena telah menunjukkan dunia luar di balik tembok pulau paradisi terima kasih eren karena telah menunjukkan apa yang ada di balik tembok karena telah menunjukkan kepadaku pemandangan ini kata armin kepada eren masih di momen eren dan armin mengobrol dalam memori obrolan sepasang sahabat itu terlihat selalu berubah latar dan waktu dimulai dari ketika mereka masih bocah hingga remaja di momen terakhir keduanya berbincang di tanah hasil injakan ribuan titan yang menewaskan manusia tanpa pandang bulu ini terlihat dalam versi manga namun versi anime menunjukkan pemandangan yang sedikit berbeda yaitu hasil injakan kaki titan itu digenangi darah manusia yang menjadi korban eren eren kemudian mengambil segenggam rambut dan tulang manusia ia tampak menyesali perbuatannya sedangkan armin mengambil kerang yang kemudian diberikan kepada eren dan berjanji untuk menyusul eren di neraka lanjut ke sebelah'