

Implementasi *Web Scraping* dan *Sentiment Analysis* Terhadap Berita Menggunakan *Machine Learning*

Ary Suryadi¹, Wahid Andika Syb'an², Nazzala Alfa'inna³, Eni Heni Hermaliani^{4*}

^{1, 2, 3, 4}Universitas Nusa Mandiri

e-mail: ¹ary.suryadi@gmail.com, ²wahid.andika176@gmail.com, ³nazzala077@gmail.com, ⁴enie_h@nusamandiri.ac.id*

*correspondence author

Abstrak

Seiring dengan perkembangan teknologi informasi yang maju seperti sekarang ini, membuat pengguna semakin mudah dalam mencari informasi melalui media *online*. Dengan banyaknya media *online* yang dapat diakses oleh masyarakat maka banyak opini yang terbentuk di masyarakat tersebar di dalamnya. Penelitian ini memiliki tujuan adalah membangun sebuah aplikasi untuk mengambil berita dari banyak media *online* dan mengeluarkan *sentiment* dari teks berita tersebut lalu membuat model *machine learning* yang mampu secara otomatis melakukan klasifikasi berita dengan benar. Metode yang digunakan pada aplikasi ini menggunakan metode *web scraping* dan model *machine learning* untuk *sentiment analysis*. Adapun data yang digunakan adalah data berita terkait dengan "Universitas Nusa Mandiri" pada bulan Desember 2022 terjaring 100 berita dari *google news* yang telah diproses oleh model *machine learning sentiment analysis* dan memberikan hasil 87% berita *sentiment* positif dan 13% berita *sentiment* negatif.

Kata Kunci: Universitas Nusa Mandiri, *Web Scraping*, *Sentiment Analysis*

Abstract

The development of advanced information technology as it is today makes it easier for users to find information through online media. With so much online media that the public can access, many opinions formed in the community spread through it. This research aims to build an application to retrieve news from many online media, extract sentiment from the news text, and then create a machine-learning model to classify news automatically. The method used in this application uses the web scraping method and machine learning models for sentiment analysis. The data is related to "Nusa Mandiri University" in December 2022; 100 news items from Google News were captured, processed by the machine-learning sentiment analysis model, and yielded 87% positive sentiment news and 13% negative sentiment news.

Keywords: Nusa Mandiri University, *Web Scraping*, *Sentiment Analysis*

1. Pendahuluan

Media *online* menurut Romli dalam bukunya yang berjudul *Jurnalistik Online* disebut juga *cyber media* (siber media), internet media (media internet), dan new media (media baru) dapat diartikan sebagai media yang disajikan secara *online* di situs *Web internet*. (Khalid, 2019)

Web scraping adalah proses pengambilan sebuah dokumen semi-terstruktur dari internet, umumnya berupa halaman-halaman *Web* dalam bahasa *markup* seperti *HTML* atau *XHTML*, dan menganalisis dokumen tersebut untuk

diambil data tertentu dari halaman tersebut untuk digunakan bagi kepentingan lain, serta banyak penelitian yang menggunakan *tools scraping* untuk mengumpulkan datanya dari *Web* (Deviacita et al., 2019)

Seiring dengan perkembangan teknologi informasi yang maju seperti sekarang ini, membuat pengguna semakin mudah dalam mencari informasi melalui media *online*. Dengan banyaknya media *online* yang bisa diakses oleh masyarakat maka opini yang terbentuk di masyarakat tersebar di media *online*. Dengan adanya *web scraping*, maka mengelola data akan

lebih efektif, hanya dengan menggunakan *library* yang tersedia pada bahasa pemrograman *Python*.

Namun saat ini masih terdapat banyak media online yang diakses masyarakat tetapi belum adanya aplikasi yang dapat mengelola berita dari banyak media online. Demikian pula belum banyak digali terkait opini atau pendapat masyarakat di media online akan menghasilkan *sentiment* negatif atau positif. Berdasarkan identifikasi masalah diatas maka rumusan masalahnya adalah (a). Bagaimana menggunakan *web scraping* yang dapat menarik berita dari banyak media *online?*, (b). Bagaimana menggunakan metode *machine learning* yang sesuai dengan *sentiment analysis*?

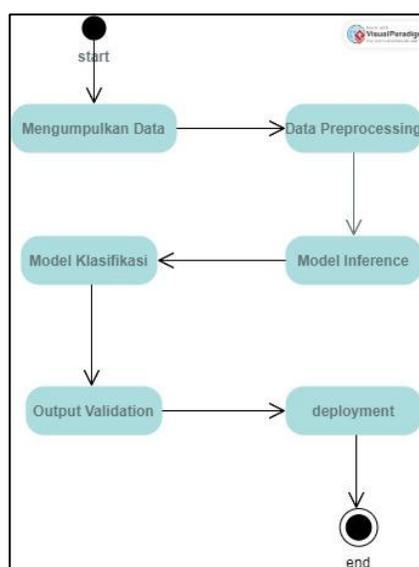
Penelitian ini menggunakan dua *sentiment analysis* yaitu, positif dan negatif. *Sentiment analysis* adalah suatu teknik *natural language processing* yang digunakan untuk menentukan status suatu data yang mengandung muatan positif, atau justru negatif. Caranya adalah dengan menggali konteks yang masih berhubungan dengan materi sumber.

Penelitian yang dilakukan oleh Hikmawan adalah adanya upaya untuk mengetahui pendapat publik dilakukan berupa sentimen positif, negatif dan netral di media twitter terhadap tindakan pencegahan Covid-19 oleh Pemerintahan di Indonesia dapat menjadi masukan bagi indikator keberhasilan pemerintah. Dari hasil pengujian menunjukkan, *accuracy* dari *Naive Bayes* sebesar 84.58%, *Support Vector Machine* sebesar 92.93%, dan *K-NN* sebesar 83.70%. Hasil *precision* dari *Naive Bayes* sebesar 82.14%, *SVM* sebesar 95.70% dan *K-NN* sebesar 80.66%. Hasil *recall* dari *Naive Bayes* sebesar 85.82%, *SVM* sebesar 89.17%, dan *K-NN* sebesar 84.13%. Dapat disimpulkan bahwa *Support Vector Machine* yang terbaik karena memiliki akurasi dan presisi tertinggi” (Hikmawan et al., 2020). Sedangkan yang dilakukan oleh Flores bahwa *Web Scraping* dengan menggunakan algoritma pencarian *Depth First Search* sampai di tingkat atau level ke-3 dapat menghasilkan hasil pencarian artikel ilmiah yang relevan dengan nilai presisi yang mencapai 0.87 dan nilai *recall* yang mencapai 0.99” (Flores et al., 2020). Lain halnya penelitian yang dilakukan Fauziyyah dimana *streaming data tweet* dengan pencarian data *collection* dimulai dari 6 April

2020 dan dari hasil analisa sentimen baik pencarian variabel COVID-19 atau *coronavirus*, nilai polaritas analisis sentimen kategori netral paling tinggi yaitu 58,94% untuk sentimen COVID-19 dan 55,10% untuk variabel *coronavirus* dibandingkan polaritas negatif atau positif” (Fauziyyah, 2020). Berdasarkan gambaran penjelasan diatas, penelitian ini mengangkat tema yaitu terkait implementasi *web scraping* dan *sentiment analysis* pada berita Universitas Nusa Mandiri menggunakan *machine learning*”.

2. Metode Penelitian

Metode penelitian yang akan dilakukan meliputi pengumpulan data dengan mengambil data sekunder yang berasal dari *Googlenews* dengan menggunakan kata kunci “Universitas Nusa Mandiri” yang di proses oleh model *machine learning sentiment analysis*. Kemudian teknik pengumpulan data yang dilakukan adalah dengan observasi mengamati hasil data sekunder yang telah diperoleh, lalu studi pustaka dari berbagai hasil penelitian sebelumnya dari paper jurnal melalui *open journal system* pada *internet*. Selanjutnya menerapkan model *Cross-Industry Standard Process for Data Mining (CRISP-DM)* yang telah disesuaikan dengan kebutuhan penelitian. *CRISP-DM* terdiri dari beberapa tahapan yaitu: *Business Understanding, Data Understanding, Data Preparation, Modeling, Evaluation, Deployment*. Tahapan penelitian terurai pada gambar 1 sebagai berikut.



Gambar 1. Diagram Tahapan Penelitian

Sumber: Suryadi, dkk (2023)

2.1. Pengertian *Python*

Python adalah bahasa pemrograman interpretatif dengan banyak aplikasi yang berfokus pada keterbacaan kode. Faktanya, *Python* juga diklaim sebagai bahasa dengan sintaks kode yang jelas, standar fungsional, dan pustaka yang komprehensif selain menggabungkan kemampuan. (Sakti, 2014)

Python memiliki tipe sistem dinamis dan manajemen memori otomatis. *Python* mendukung beberapa *paradigma* pemrograman, diantaranya pemrograman berorientasi objek, imperatif, fungsional dan prosedural, serta memiliki *library* yang besar dan komprehensif. (Suprpto dan Ni'mah, 2018)

Python memiliki sejumlah karakteristik, salah satunya adalah bahasa pemrograman dinamis dengan manajemen memori otomatis bawaan. *Python* biasanya digunakan sebagai bahasa *scripting*, namun dalam penggunaan sebenarnya, *python* juga digunakan dalam situasi di mana ia biasanya tidak digunakan dalam bahasa pemrograman. *Python* bisa digunakan untuk berbagai kebutuhan untuk pengembangan perangkat lunak dan dapat berjalan pada berbagai *platform* sistem operasi. Beberapa diantaranya adalah sebagai berikut: *Windows, Linux/Unix, OS/2, Mac OS X, Amiga, Palm, Symbian, Java Virtual Machine*. (Komputermedia, 2020)

2.2. Pengertian *Sentiment Analysis*

Sentiment Analysis atau analisis sentimen adalah proses memahami dan mengelompokkan perasaan positif, negatif, dan netral yang diungkapkan secara tertulis. Penambahan opini adalah nama lain untuk *Sentiment Analysis*. *Sentiment Analysis* menghemat waktu dan tenaga Anda, yang merupakan salah satu manfaatnya. Kini, sentimen Analisis dapat dilakukan secara otomatis. (Arviana, 2021)

2.3. Pengertian *Naive Bayes*

Menurut Olson Delen (2008), menjelaskan *Naive Bayes* untuk setiap kelas keputusan, menghitung probabilitas dengan syarat bahwa kelas keputusan adalah benar, mengingat *vektor* informasi obyek. Algoritma ini mengasumsikan bahwa atribut obyek adalah independen. (Widiyanto, 2020)

2.4. Pengertian *Machine Learning*

Secara definisi, *Machine learning* merupakan ilmu atau studi yang mempelajari tentang algoritma dan model statistik yang digunakan oleh sistem komputer untuk melakukan *task* tertentu tanpa instruksi eksplisit. *Machine learning* bergantung pada pola dan kesimpulan (Dinata & Hasdyna, 2020)

Salah satu cabang dari kecerdasan buatan adalah *machine learning*, yang berfokus pada studi tentang bagaimana komputer belajar dari data untuk meningkatkan kecerdasan mereka. (Wahyono, 2018)

2.5. Pengertian *Data Mining*

Menurut Sudirman, Windarto dan Wanto (2018), *data mining* merupakan proses yang menggunakan berbagai teknik dan alat analisa data untuk menemukan hubungan dan pola yang tersembunyi. Pendekatan dasar dalam *data mining* adalah untuk meringkas data dan untuk mengekstrak informasi berguna yang masuk akal dan sebelumnya tidak diketahui. (Wanto, 2020)

Data mining adalah suatu proses menganalisis pola data yang tersembunyi menurut berbagai perspektif untuk kategorisasi menjadi informasi yang berguna, yang dikumpulkan di area umum, *data warehouse* untuk analisa, dan informasi lainnya. (Arhami & Nasir, 2020)

2.6. Pengertian *Cross-Industry Standard Process for Data Mining (CRISP-DM)*

Mariscal, Marba dan Fernandez (Mariscal, Marban, and Fernandez 2010) menyatakan *CRISP-DM* sebagai *de facto* menjadi standar untuk pengembangan proyek data mining dan *knowledge discovery* karena paling banyak digunakan dalam pengembangan data *mining*. (Mauritsius & Binsar, 2020)

2.7. Pengertian *Web Scraping*

Web scraping dapat digunakan untuk mencari tahu tentang pesaing perusahaan, melakukan riset pasar, membuatnya lebih mudah untuk mendapatkan prospek, dan mendapatkan informasi spesifik dari seluruh halaman. Salah satu pemanfaatan *Web Scraping* adalah mengkopir daftar kontak dari sebuah direktori *Website*, atau bisa digunakan untuk merekap *e-mail* pelanggan yang melakukan subscribe. (Pradana, 2021)

3. Hasil dan Pembahasan

3.1. Business Understanding

Pada tahapan *Business Understanding*, dilakukan pemahaman terhadap objek penelitian, pemahaman objek penelitian dilakukan dengan mengambil data dari berbagai situs berita dengan menggunakan kata kunci "Universitas Nusa Mandiri" yang berbahasa Indonesia. Pada fase ini data dari berbagai situs berita tersebut akan diambil menggunakan teknik *web scraping* dengan menggunakan bahasa pemrograman *Python version 3*.

```

from pygsheets import GoogleSheets
gn = GoogleNews()
gn = GoogleNews(lang = 'id', country = 'ID')
um = ['Universitas+NusaMandiri']
keywords = um

search_results = []
total_berita = 0
print("Mencari Berita-berita dengan keyword "Universitas Nusa Mandiri" dari tanggal 1 Des 2022 s/d 31 Des 2022")
for topic in keywords:
    search = gn.search(topic, from_='2022-12-01', to_='2022-12-31')
    result = search["entries"]

    if (len(result) > 0):
        total_berita = total_berita + len(result)
        print("Ditemukan:", total_berita, "Link Berita")

print("Selesai.")
    
```

Gambar 2. Teknik *Web Scraping* dengan *Python*
Sumber: Suryadi, dkk (2023)

3.2. Data Understanding

Pada tahap ini dilakukan proses memahami data yang diperoleh dari banyak situs berita dengan menggunakan kata kunci "Universitas Nusa Mandiri" yang berbahasa Indonesia. Sebanyak 100 data yang diambil dalam rentang waktu dari tanggal 1 Desember 2022 sampai dengan tanggal 31 Desember 2022. Dari berbagai macam situs berita sebanyak 100 data, hasil dari tahap ini adalah sebagai berikut:

	published	situs	judul	link
0	Tue, 20 Dec 2022 08:00:00 GMT	Republika Online	Universitas Nusa Mandiri Hadiri Pameran Persef...	https://news.google.com_/news/articlesCB...
1	Wed, 07 Dec 2022 08:00:00 GMT	Republika Online	Melahirkan UIN Kupera Jember dengan Proklamasi S...	https://news.google.com_/news/articlesCB...
2	Wed, 28 Dec 2022 09:28:43 GMT	Republika Online	Kampus Digital Bisnis UINM Sukses Gelar Webinar...	https://news.google.com_/news/articlesCB...
3	Tue, 20 Dec 2022 08:00:00 GMT	Republika Online	Jadi Rekan di Yayasan Pendidikan Bina, Iki...	https://news.google.com_/news/articlesCB...
4	Tue, 20 Dec 2022 08:00:00 GMT	Republika Online	Kampus Digital Bisnis UINM Fasilitas Magang S...	https://news.google.com_/news/articlesCB...
...
95	Fri, 02 Dec 2022 08:00:00 GMT	Por-Kupang.com	Bank NTT Raih Tiga Penghargaan Dari Bank Indon...	https://news.google.com_/news/articlesCB...
96	Sun, 04 Dec 2022 08:00:00 GMT	Republika Online	Rupati Banyuwangi Cantangkan Gerakan Nasional Ser...	https://news.google.com_/news/articlesCB...
97	Sun, 04 Dec 2022 08:00:00 GMT	Por-Kupang.com	DRMB Lulus Hibah Dana Jombang Berupa Saat Kom...	https://news.google.com_/news/articlesCB...
98	Sun, 04 Dec 2022 08:00:00 GMT	radobukatum.com	Penyetas Gempa Cianjur Dapat Uang Sewa Rumah...	https://news.google.com_/news/articlesCB...
99	Sun, 04 Dec 2022 08:00:00 GMT	Republika Online	Perilaku APD Semis, Cegahwaji 1979 Jua Manj...	https://news.google.com_/news/articlesCB...

Gambar 3. Hasil Proses Pemahaman Data
Sumber: Suryadi, dkk (2023)

3.3. Data Preparation

Pada tahap ini yaitu melakukan ekstraksi *metadata* pada *dataset* berita-berita yang sudah dikumpulkan pada tahap

sebelumnya, hal ini dilakukan untuk mengambil bagian-bagian berita yang diperlukan untuk penelitian ini. Selanjutnya *dataset* tersebut akan diklasifikasikan menggunakan model *machine learning* dan *preprocessing* yang sudah ada dalam bentuk model *package (pickle file)* sehingga akan menghasilkan klasifikasi sentimen yaitu positif atau negatif.



Gambar 4. Diagram *Data Preparation*
Sumber: Suryadi, dkk (2023)

3.4. Modeling

Model *machine learning* yang akan digunakan pada penelitian ini menggunakan model yang saat ini sudah ada dalam bentuk model *package (pickle file)* sehingga tidak perlu dilakukan proses model *training*. Penelitian ini tidak membahas mengenai bagaimana cara pembuatan model *machine learning* untuk melakukan klasifikasi sentimen analisis, tapi berfokus pada bagaimana cara menggunakan model yang sudah ada.



Gambar 5. Diagram *Modeling*
Sumber: Suryadi, dkk (2023)

3.5. Evaluation

Pada tahap ini menggunakan teknik *web scraping* dengan menggunakan bahasa pemrograman *python*. Yang dimana hasil dari *web scraping* dengan objek penelitian "Universitas Nusa Mandiri" yang diambil dari berbagai situs berita akan diklasifikasikan kedalam model *machine learning* guna untuk mendapatkan hasil sentimen positif atau negatif.

```

import re
import pandas as pd
import requests
from bs4 import BeautifulSoup
from sklearn.feature_extraction.text import TfidfVectorizer
from sklearn.naive_bayes import MultinomialNB
from sklearn.pipeline import Pipeline
from sklearn.metrics import accuracy_score

# URL of the website to scrape
url = "https://www.unsma.ac.id/"

# Send a GET request to the website
response = requests.get(url)

# Parse the HTML content
soup = BeautifulSoup(response.text, 'html.parser')

# Find all news articles
articles = soup.find_all('div', class_='news-item')

# Extract the title and text of each article
titles = []
texts = []

for article in articles:
    title = article.find('h3').text
    text = article.find('p').text

    titles.append(title)
    texts.append(text)

# Create a DataFrame
df = pd.DataFrame({'title': titles, 'text': texts})

# Preprocess the text
vectorizer = TfidfVectorizer()
X = vectorizer.fit_transform(df['text'])

# Train a Naive Bayes classifier
model = MultinomialNB()
model.fit(X, df['title'])

# Predict the sentiment of the news articles
y_pred = model.predict(X)

# Evaluate the model
accuracy = accuracy_score(y_pred, df['title'])
print("Accuracy: ", accuracy)
    
```

Gambar 6. Kode *Web scraping Python*
Sumber: Suryadi, dkk (2023)

Kode pemrograman *Web scraping Python* bertujuan untuk mengambil data

berita dari berbagai situs dengan objek penelitian "Universitas Nusa Mandiri".

```

14 def data_preprocessing(data):
15     import pickle
16     import numpy as np
17     import re
18     import string
19     from sklearn.feature_extraction.text import TfidfVectorizer
20     from sklearn.feature_extraction.text import StopWordRemoverFactory
21     from Sastrawi.Stemmer.StemmerFactory import StemmerFactory
22
23     model = pickle.load(open('model_sentiment.pkl', 'rb'))
24     vec = pickle.load(open('final_count_vectorizer.pkl', 'rb'))
25
26     def model_inference(kalimat):
27         # Preprocessing Data
28         textlowercase = kalimat.lower()
29         textremovnumber = ''.join(item for item in textlowercase if not item.isdigit())
30         textremovepunc = re.sub('[^a-zA-Z]', '', textremovnumber)
31         textgabung = np.array([textremovepunc])
32         textpunct = ''.join(textgabung)
33         textpunct = textpunct.translate(str.maketrans('', '', string.punctuation))
34         factorystem = StemmerFactory()
35         stemmer = factorystem.create_stemmer()
36         textstemming = stemmer.stem(textpunct)
37         factorystop = StopWordRemoverFactory()
38         stopword = factorystop.create_stop_word_remover()
39         textstopword = stopword.remove(textstemming)
40
41         # Transform Vec & TFIDF Process
42         trans_vec = vec.transform([textstopword])
43
44         # Model Prediction
45         predict_result = model.predict(trans_vec)
46         if (predict_result == 1):
47             result = "Positif"
48         else:
49             result = "Negatif"
50
51         return result
52
53 # Load Dataset
54 data['sentiment'] = data.apply(lambda x: model_inference(x))
55 value = data.to_dict('records')
56 return value

```

Gambar 7. Kode data *Preprocessing* dan *modeling sentiment Python*
Sumber: Suryadi, dkk (2023)

Kode pemrograman *preprocessing* ini bertujuan untuk melakukan tahap *text preprocessing* pada berita atau informasi yang diambil dari tahap sebelumnya. *Preprocessing* teknik yang digunakan adalah:

1. *Lower Cases*: Mengubah semua kata didalam kalimat menjadi huruf kecil.
2. *Remove Number*: Menghilangkan angka didalam kalimat.
3. *Tokenize*: Mengubah kalimat kedalam bentuk token.
4. *Remove Punctuation*: Menghilangkan tanda baca didalam kalimat.
5. *Stemming*: Mengambil kata baku didalam kalimat.
6. *Stopword Removal*: Menghilangkan kata sambung didalam kalimat.

Setelah tahap *text-preprocessing* selesai dilanjutkan ke dalam tahap sentimen analisis ke dalam model yang sudah ada. Dengan menggunakan teknik *count vectorize*. Hal ini bertujuan untuk mendapatkan hasil sentimen positif atau negatif dari suatu berita yang sudah didapat.

3.6. Deployment

Pada tahap ini membuat sebuah *application platform* menggunakan bahasa pemrograman *Python 3*. Berikut adalah tahapan membuat *application platform*:

1. Instalasi *Python Framework FastAPI*

```

bash
$ pip install fastapi

```

Gambar 8. *Install FastAPI*
Sumber: Suryadi, dkk (2023)

2. Instalasi *Python Web Server*

```

bash
$ pip install "uvicorn[standard]"

```

Gambar 9. *Install Uvicorn*
Sumber: Suryadi, dkk (2023)

Tahap kedua, pada proses selanjutnya adalah melakukan instalasi *Uvicorn* yang berfungsi untuk *handling request* ke *application platform* pada tahap ini.

3. Menjalankan *Web Server*

```

bash
$ uvicorn main:app --reload

INFO: Uvicorn running on http://127.0.0.1:8000 (Press CTRL+C to quit)
INFO: Started reloader process [28720]
INFO: Started server process [28722]
INFO: Waiting for application startup.
INFO: Application startup complete.

```

Gambar 10. Menjalankan *Web Server*
Sumber: Suryadi, dkk (2023)

Tahap ketiga, menjalankan *Web server* dengan *port default 8000*, sehingga *application platform* ini bisa diakses melalui url `http://127.0.0.1:8000`.

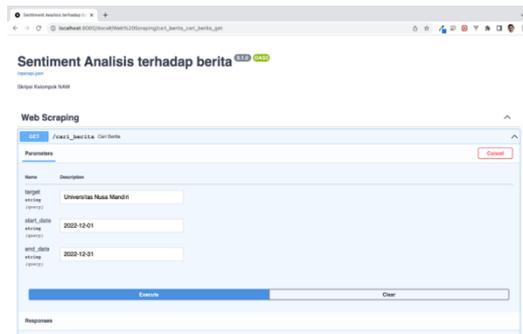
4. Tampilan *Application platform*

Berikut adalah tampilan *Application platform*:

Pada kolom 1: kolom tersebut untuk mengisi *keyword* yang ingin dicari.

Pada Kolom 2: kolom untuk mengisi tanggal awal *keyword* yang ingin dicari.

Pada kolom 3: kolom untuk mengisi tanggal akhir *keyword* yang ingin dicari.



Gambar 11. Tampilan *application platform*
Sumber: Suryadi, dkk (2023)

5. Output hasil sentimen analisis pada dataset berita dalam bentuk .csv file

publikasi	Waktu	Judul	Situs	Sentiment
Tan, 20 Dec 2022 08:00:00 GMT	Republika Online	Universitas Nusa Mandiri Hadiri Pameran Pendidikan Aera Festival 2022	https://www.google.com/_/v/.../v/.../v/.../v/.../v/.../v/.../...	sentiment
Wah, 28 Dec 2022 08:00:00 GMT	Republika Online	Kampus Digital Bisnis UHM Sukses Gelar Webinar Visiting & Sharing Univer	https://www.google.com/_/v/.../v/.../v/.../v/.../v/.../v/.../...	positif
Thu, 20 Dec 2022 08:00:00 GMT	Republika Online	Kampus Digital Bisnis UHM Festivali Mengajar Siswa UHM, Regerasi di Sukare	https://www.google.com/_/v/.../v/.../v/.../v/.../v/.../v/.../...	positif
Man, 19 Dec 2022 08:00:00 GMT	Republika Online	UHM akan Hadiri di Edufair MUBK SMA Kota Bekasi 2022	https://www.google.com/_/v/.../v/.../v/.../v/.../v/.../v/.../...	positif
Thu, 15 Dec 2022 08:00:00 GMT	Republika Online	Mahasiswa Kampus Digital Bisnis UHM Pak-Prekara di Komputasi IoT	https://www.google.com/_/v/.../v/.../v/.../v/.../v/.../v/.../...	positif
Man, 19 Dec 2022 08:00:00 GMT	Republika Online	UHM Prati Sistem Informasi Gelar Workshop Interactive Design UJU/UK Post	https://www.google.com/_/v/.../v/.../v/.../v/.../v/.../v/.../...	positif
Thu, 15 Dec 2022 08:00:00 GMT	Republika Online	Prati Manajemen Kampus UHM Sukses Gelar Webinar Perencanaan Keumaha	https://www.google.com/_/v/.../v/.../v/.../v/.../v/.../v/.../...	positif
Thu, 29 Dec 2022 08:00:00 GMT	Republika Online	Universitas BSI Bekali Ilmu Video Editing Pada Program PHM PKM 2022	https://www.google.com/_/v/.../v/.../v/.../v/.../v/.../v/.../...	positif
Thu, 14 Dec 2022 08:00:00 GMT	Republika Online	Universitas BSI Bekali Ilmu Fotografi Bagi Mahasiswa Program PHM PKM	https://www.google.com/_/v/.../v/.../v/.../v/.../v/.../v/.../...	positif
Wah, 21 Dec 2022 08:00:00 GMT	Republika Online	Prati Ilmu Komputer Kampus Digital Bisnis UHM Sukses Gelar Webinar Pytho	https://www.google.com/_/v/.../v/.../v/.../v/.../v/.../v/.../...	positif
Man, 19 Dec 2022 08:00:00 GMT	Kompasiana.com	Penetapan Microsoft Power Point pada Pengrus dan Staf TPQ Bina Ummah	https://www.google.com/_/v/.../v/.../v/.../v/.../v/.../v/.../...	positif
Sat, 24 Dec 2022 08:00:00 GMT	Kompasiana.com	Obser: dan Mahasiswa Melakukan Pengabdian Masyarakat	https://www.google.com/_/v/.../v/.../v/.../v/.../v/.../v/.../...	positif
Sat, 24 Dec 2022 08:00:00 GMT	ANTARA	ANTARA Menggaji PPKUM Universitas Pancasila gelar seminar akademi	https://www.google.com/_/v/.../v/.../v/.../v/.../v/.../v/.../...	positif
Wah, 28 Dec 2022 18:28:00 GMT	Kader Berita	Rakasa Daging Berleka Aera Festival 2022 SIKMA 4 Daging	https://www.google.com/_/v/.../v/.../v/.../v/.../v/.../v/.../...	positif
Tue, 27 Dec 2022 08:00:00 GMT	ANTARA	Bani Mandiri perkenalkan mahasiswa bertamabaki lewat KCP Unsur	https://www.google.com/_/v/.../v/.../v/.../v/.../v/.../v/.../...	positif
Tue, 20 Dec 2022 08:00:00 GMT	Republika Online	Pendaftaran dan Regulasi Akad Perawatan UHM Dukung KWT	https://www.google.com/_/v/.../v/.../v/.../v/.../v/.../v/.../...	positif
Sat, 08 Dec 2022 08:00:00 GMT	Republika Online	33 - Republika Online Kampus Digital Bisnis UHM Sukses Lunak 1.714 Mahasiswa Pada Prosa	https://www.google.com/_/v/.../v/.../v/.../v/.../v/.../v/.../...	positif
Man, 12 Dec 2022 08:00:00 GMT	Kompas.com	1000 Cara Daftar 4000 Kuliah untuk SMPN 2022, Calon Mahasiswa Simak	https://www.google.com/_/v/.../v/.../v/.../v/.../v/.../v/.../...	positif
Wed, 07 Dec 2022 08:00:00 GMT	Republika Online	Mahasiswa UHM Kampus Jabataringin Promosikan Bisnis di Entrepreneur	https://www.google.com/_/v/.../v/.../v/.../v/.../v/.../v/.../...	positif
Thu, 29 Dec 2022 08:00:00 GMT	Republika Online	172 - Universitas Islam Gelar Webinar	https://www.google.com/_/v/.../v/.../v/.../v/.../v/.../v/.../...	positif
Sat, 08 Dec 2022 08:00:00 GMT	Republika Online	Hempunan Mahasiswa Prati Perforansi UHM Galang Dana Prati Gemasa	https://www.google.com/_/v/.../v/.../v/.../v/.../v/.../v/.../...	positif
Tue, 20 Dec 2022 08:00:00 GMT	ANTARA Bat	Kepri Bat jabaikan periksa 20 juta keas dalam SP Unud	https://www.google.com/_/v/.../v/.../v/.../v/.../v/.../v/.../...	negatif
Fri, 09 Dec 2022 08:00:00 GMT	Republika Online	Kisah Anak Panti yang Berhasil Kuliah Dengan Beasiswa di UHM	https://www.google.com/_/v/.../v/.../v/.../v/.../v/.../v/.../...	positif
Thu, 08 Dec 2022 08:00:00 GMT	Republika Online	Prati Ilmu Komputer Kampus UHM Sukses Gelar Webinar Pengantun Publik	https://www.google.com/_/v/.../v/.../v/.../v/.../v/.../v/.../...	positif
Fri, 16 Dec 2022 08:00:00 GMT	Republika Online	Pemangaran Gelar Bekas Uniba Singaperbangsa, Prof Laminan	https://www.google.com/_/v/.../v/.../v/.../v/.../v/.../v/.../...	positif
Thu, 08 Dec 2022 08:00:00 GMT	Republika Online	Raih 4 Gelar Sains, Diriga Tetap Kuliah di Kampus Digital Bisnis UHM	https://www.google.com/_/v/.../v/.../v/.../v/.../v/.../v/.../...	positif
Thu, 22 Dec 2022 08:00:00 GMT	Republika Online	Galasi Komptar Keluarin Prosa Keasara	https://www.google.com/_/v/.../v/.../v/.../v/.../v/.../v/.../...	negatif
Man, 12 Dec 2022 08:00:00 GMT	UNH Dampiang	1 Kurang Pencemaran Limbah Cair Tahu, TBM PKM	https://www.google.com/_/v/.../v/.../v/.../v/.../v/.../v/.../...	positif

Gambar 12. Hasil Sentimen pada kolom
"sentiment"
Sumber: Suryadi, dkk (2023)

Output dari *application platform* ini berbentuk *json file*, sehingga mudah diubah kedalam *format file* lain, misalnya *format csv file*.

Penelitian ini membuat aplikasi dan menerapkan algoritma *machine learning* untuk mengolah *sentiment analysis* menggunakan model algoritma *Naïve Bayes*. Aplikasi dapat mengambil berita dengan cepat dan mengklasifikasikan berita positif dan negatif pada media online, sehingga memudahkan pengguna mengetahui sentimen dari berita di media online terhadap berita Universitas Nusa Mandiri.

4. Kesimpulan

Berdasarkan hasil penelitian ini, untuk melakukan *web scraping* agar dapat menarik berita dari berbagai media *online* dengan cara menggunakan *python library pygooglenews*, yang dimana tekniknya dengan cara mengirimkan *HTTP request* kepada *server google* menggunakan kata kunci dan rentang tanggal serta *region/country* sebagai parameter untuk mendapatkan berita yang ingin diperoleh. Implementasi *web scraping* pada penelitian ini hanya melibatkan mesin pencari *Google News* dengan kata kunci "Universitas Nusa Mandiri". Adapun metode *sentiment analysis* yang digunakan adalah *Aspect-Based Sentiment Analysis*. Kemudian hasil olah data diperoleh dengan kata kunci "Universitas Nusa Mandiri" dari rentang tanggal 01 Desember 2022 sampai dengan 31 Desember 2022 mendapatkan 100 berita dengan hasil 87% berita sentimen positif dan 13% berita sentimen negatif.

Referensi

Arhami, M., & Nasir, M. (2020). *Data Mining - Algoritma dan Implementasi - Google Buku*.

<https://books.google.co.id/books?id=AtCEAAQBAJ&printsec=frontcover&hl=id#v=onepage&q&f=false>

Arviana, G. N. (2021, February 1). *Sentiment Analysis: Pengertian, Teknik, dan Penggunaannya - Glints Blog*. Glints.Com.

<https://glints.com/id/lowongan/sentiment-analysis/#.Y4c35hRBzrc>

Deviacita, D., #1, A., Sasty, H., #2, P., Muhardi, H., Profesor, J., Nawawi, D. H. H., Laut, B., Tenggara, P., Pontianak, K., & Barat, K. (2019). *Implementasi Web Scraping untuk Pengambilan Data pada Situs Marketplace*. 7(4).

Dinata, R. K., & Hasdyna, N. (2020). *e-book Machine Learning*.

Fauziyyah, A. K. (2020). ANALISIS SENTIMEN PANDEMI COVID19 PADA STREAMING TWITTER DENGAN TEXT MINING PYTHON. *Jurnal Ilmiah SINUS*, 18(2), 31. <https://doi.org/10.30646/sinus.v18i2.491>

- Flores, V. A., Permatasari, P. A., & Jasa, L. (2020). Penerapan Web Scraping Sebagai Media Pencarian dan Menyimpan Artikel Ilmiah Secara Otomatis Berdasarkan Keyword. *Majalah Ilmiah Teknologi Elektro*, 19(2), 157. <https://doi.org/10.24843/mite.2020.v19i02.p06>
- Hikmawan, S., Pardamean, A., Nur Khasanah, S., Mandiri, N., Damai No, J., Jati Barat, W., & Selatan, J. (2020). Halaman: 167-176 Terakreditasi Peringkat 5 (SINTA 5) sesuai SK RISTEKDIKTI Nomor (Vol. 20, Issue 2). <http://ejurnal.ubharajaya.ac.id/index.php/JKI>
- Khalid, I. (2019). KREDIBILITAS MEDIA CETAK DAN MEDIA ONLINE. *Penelitian Sosial Keagamaan*. www.ejournal.annadwahkualatungkal.ac.id
- Komputermedia. (2020, January 24). *Pengertian Python, Fungsi, Sejarah & Belajar Pemrograman Python*. <https://www.komputermedia.com/pengertian-python/>
- Mauritsius, T., & Binsar, F. (2020, September 18). *Cross-Industry Standard Process for Data Mining (CRISP-DM) – MMSI BINUS University*. <https://mmsi.binus.ac.id/2020/09/18/cross-industry-standard-process-for-data-mining-crisp-dm/>
- Pradana, G. (2021, February 15). *Web Scraping Pengertian, Teknik, Manfaat dan Kendala adalah*. <https://ngalup.co/articles/pengertian-teknik-manfaat-kendala-web-scraping/>
- Sakti. (2014). *Belajar Python*. <http://sakti.github.io/>
- Suprpto dan Ni'mah, Y. L. (2018). *Pengantar Analisa Data Menggunakan Python*. <https://deepublishstore.com/shop/buku-pengantar-analisis-3/>
- Wahyono, T. (2018). *Fundamental of Python for Machine Learning: Dasar-Dasar Pemrograman Python untuk Machine Learning dan Kecerdasan Buatan*. <https://www.researchgate.net/publication/330441937>
- Wanto, A. (2020). *Data Mining : Algoritma dan Implementasi - Anjar Wanto - Google* Buku. <https://books.google.co.id/books?id=LSPfDwAAQBAJ&printsec=frontcover&hl=id#v=onepage&q&f=false>
- Widianto, M. H. (2020, July). *Algoritma Naive Bayes | BINUS UNIVERSITY BANDUNG - Kampus Teknologi Kreatif*. <https://binus.ac.id/bandung/2019/12/algoritma-naive-bayes/>