

# Analisis Sentimen Migrasi Televisi Digital pada Twitter Menggunakan Perbandingan Algoritma *Multinomial Naïve Bayes*, *Support Vector Machines*, dan *Logistic Regression*

Ryo Benhard Dahlian<sup>[1]\*</sup>, Delima Sitanggang<sup>[2]</sup>

Fakultas Sains dan Teknologi<sup>[1],[2]</sup>

Universitas Prima Indonesia

Medan, Indonesia

ryobenhard08@gmail.com<sup>[1]</sup>, delimasitanggang@unprimdn.ac.id<sup>[2]</sup>

**Abstract**— The Ministry of Communication and Information Technology (KEMENKOMINFO) has announced to the publics in Indonesia regarding the termination of analog television broadcasts or called analog switch-off, which requires the public to migrate from analog television to digital television. Regarding the process of stopping analog broadcasts this raises pros and cons by the people in Indonesia. Many people give their respective opinions through social media, especially on Twitter. A collection of pros and cons data from the public can be collected and used as research of sentiment analysis. This research will focus on comparing three classification algorithms, which is called Multinomial Naïve Bayes, Support Vector Machines, and Logistic Regression using the same dataset and the same method called Lexicon Based. The results showed that the highest accuracy is Support Vector Machines with the accuracy is 94.00%, Logistic Regression with the accuracy is 90.00%, and Multinomial Naïve Bayes with the accuracy is 88.00%.

**Keywords**— Sentiment analysis, Multinomial Naïve bayes, Support Vector Machines, Logistic regression, Twitter

**Abstrak**— Pemerintah melalui Kementerian Komunikasi dan Informatika (KEMENKOMINFO) telah menginformasikan kepada masyarakat Indonesia mengenai penghentian siaran televisi analog (*analog switch-off*) yang mengharuskan masyarakat untuk melakukan migrasi dari televisi analog menuju televisi digital. Tentunya mengenai proses penghentian siaran analog ini menimbulkan pro dan kontra oleh masyarakat di Indonesia. Masyarakat banyak memberikan pendapat mereka masing-masing melalui media sosial, khususnya pada media sosial Twitter. Kumpulan data pro dan kontra dari masyarakat dapat dikumpulkan dan dijadikan penelitian mengenai analisis sentimen. Pada penelitian ini akan berfokus pada perbandingan tiga algoritma klasifikasi, yaitu *Multinomial Naïve Bayes*, *Support Vector Machines*, dan *Logistic Regression* serta menggunakan dataset dan metode yang sama yaitu *Lexicon Based*. Hasil penelitian menunjukkan bahwa tingkat akurasi terbaik adalah *Support Vector Machines* dengan akurasi sebesar 94,00%, *Logistic Regression* dengan akurasi sebesar 90,00%, dan *Multinomial Naïve Bayes* dengan akurasi sebesar 88,00%.

**Kata Kunci**— Analisis sentimen, *Multinomial Naïve Bayes*, *Support Vector Machines*, *Logistic regression*, Twitter

## I. PENDAHULUAN

Perkembangan media sosial menjadi semakin signifikan di era teknologi pada beberapa tahun terakhir. Jumlah pengguna media sosial terus meningkat pesat dari hari ke hari. Setiap pengguna media sosial dapat melakukan posting mengenai pandangan, pemikiran, hingga kisah kehidupan mereka di media sosial dengan sebebas-bebasnya tanpa ada batasan apa pun. Karena mudahnya aksesibilitas dari media sosial, pengguna mulai melakukan migrasi dari sarana komunikasi yang tradisional seperti blog atau mailing list ke situs berbasis *microblogging* yaitu: Twitter, Facebook, dan lainnya [1].

Pemerintah Indonesia melalui Kementerian Komunikasi dan Informatika telah menginformasikan kepada masyarakat diseluruh Indonesia mengenai penghentian siaran televisi analog (*analog switch-off*) yang mengharuskan masyarakat melakukan migrasi dari televisi analog menuju televisi digital sejak akhir tahun 2021. Masyarakat yang hanya memiliki televisi analog dapat mengakses siaran televisi digital dengan bantuan alat bernama *set top box* (STB) yang bersertifikasi KOMINFO. Penghentian siaran televisi analog merupakan proses peralihan siaran analog siaran digital, atau dengan kata lain merupakan proses dimana siaran televisi analog akan dihentikan dan digantikan siaran televisi digital [2]–[5].

Sehingga mengenai proses penghentian siaran analog ini menimbulkan pro dan kontra oleh masyarakat luas di Indonesia, baik mengenai proses pelaksanaan hingga ketidaksetujuan masyarakat mengenai proses ini. Melalui media sosial, banyak masyarakat mengungkapkan pandangan masing-masing tentang tindakan yang akan dilakukan pemerintah tersebut. Kumpulan data yang digunakan sebagai penelitian tentang proses penghentian siaran analog ini diambil melalui *Twitter*. Kumpulan data pro dan kontra tersebut dapat dilakukan analisis sentimen dengan memanfaatkan teknologi komputer, yaitu pemanfaatan proses penggalian data (*data mining*) [6]–[10].

Analisis sentimen (*opinion mining*) merupakan salah satu cabang ilmu penelitian dari kelompok *text mining* yang bertujuan agar dapat menentukan persepsi atau subjektivitas publik (khalayak) terhadap suatu fenomena, kejadian,

permasalahan, atau topik pembahasan dengan melakukan prediksi emosi yang terkandung di dalam sebuah kata, kalimat, atau kumpulan dokumen dengan maksud agar dapat memahami sikap, pendapat, serta emosi yang diungkapkan melalui media sosial [8][11].

Karya ilmiah yang membahas analisis sentimen telah banyak dibahas pada penelitian sebelumnya, pada penelitian menggunakan algoritma *Multinomial Naive Bayes* menghasilkan akurasi sebesar 73,86% [12] dan 69,23% [13]. Penelitian yang menggunakan algoritma *Support Vector Machines* menghasilkan akurasi sebesar 96,68% [14] dan 88,57% [15], sedangkan penelitian yang menggunakan algoritma *Logistic Regression* menghasilkan akurasi terbaik sebesar 77,00% [16] dan 87,02% [17]. Tujuan penggunaan ketiga algoritma klasifikasi tersebut dikarenakan berdasarkan penelitian terdahulu memiliki hasil yang terlihat cukup baik.

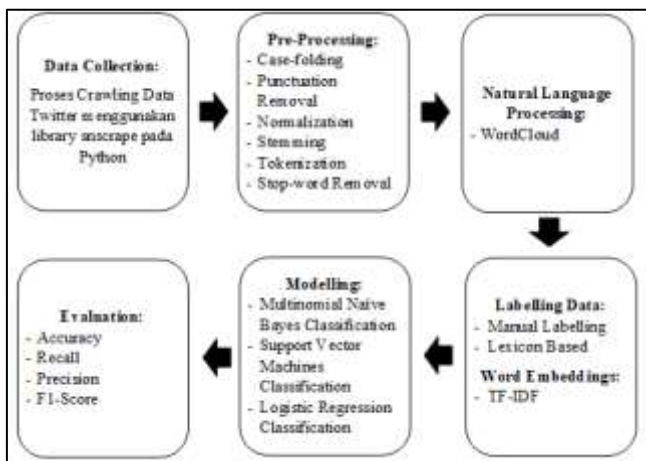
Penelitian terdahulu berfokus pada hasil *accuracy* dengan metode dan dataset yang berbeda pada penelitian. Fokus pada penelitian ini adalah membandingkan nilai *confusion matrix* (*accuracy*, *precision*, *recall*, dan *f1-score*) ketiga algoritma klasifikasi diantara *Multinomial Naive Bayes*, *Support Vector Machines*, dan *Logistic Regression* dengan menggunakan metode manual dan *Lexicon Based* dan *dataset* yang sama untuk seluruh algoritma klasifikasi, agar dapat melihat apakah dengan menggunakan metode ini akan menghasilkan nilai *accuracy* yang lebih baik dari penelitian sebelumnya, serta melihat hasil klasifikasi terbaik sehingga dapat dilakukan perbandingan dengan klasifikasi lain di penelitian selanjutnya.

## II. METODOLOGI PENELITIAN

### A. Hardware dan Software

Pada penelitian ini dibutuhkan penggunaan *hardware* dan *software* guna mendukung keberhasilan dari pelaksanaan penelitian ini. *Software* yang digunakan untuk penelitian adalah Python versi 3.7, Jupyter Notebook versi 6.4.8, serta Microsoft Excel. Sedangkan *hardware* yang digunakan dalam adalah dengan Prosesor Intel Core i7-10510U CPU @ 2.30 GHz, RAM 16 GB, dan Sistem Operasi Windows 10 Pro.

Tahapan-tahapan yang dilakukan pada penelitian ini dapat dilihat pada Gambar 1.



Gambar 1. Tahapan Penelitian

### B. Data Collection dan Data Selection

Tahapan *data collection* adalah proses untuk mengumpulkan sumber data yang akan digunakan sebagai bahan penelitian. Pada penelitian yang dilaksanakan, pengumpulan sumber data diambil melalui media sosial Twitter. Proses pengumpulan sumber data melalui Twitter dilakukan menggunakan teknik *crawling data* dengan *library sncraspe*. *Sncraspe* adalah salah satu fitur *scraper* untuk layanan jejaring sosial. Sistem akan mengambil profil pengguna, *hashtag*, tanggal dan waktu, atau konten *posting*.

Dalam proses pencarian dan pengumpulan sumber data, dilakukan pembatasan atau syarat khusus untuk sumber data yang diambil berdasarkan ketentuan berikut ini:

1. Kata kunci yang digunakan untuk penelitian adalah *keyword* yang berhubungan dengan "migrasi televisi digital"
2. Pencarian dataset yang ditentukan hanya teks dengan Bahasa Indonesia
3. Tanggal pengambilan dataset pada Twitter yang digunakan untuk penelitian direntang tanggal 31 Oktober 2022 sampai dengan 3 November 2022

Proses *data collection* menggunakan teknik *crawling data* dengan *library sncraspe* dapat dilihat pada Gambar 2.

```

#Crawling Data
pd.options.display.max_colwidth = 500

query = "(migrasi televisi digital) lang:id until:2022-11-03 since:2022-10-31"
tweets = []
limit = 100

@time
try:
    print("-----")
    print("Mulai Proses Crawling Data")
    for tweet in sntwitter.TwitterSearchScraper(query = query).get_items():
        if len(tweets) == limit:
            break
        else:
            tweets.append([tweet.date, tweet.user.username, tweet.content])
    df = pd.DataFrame(tweets, columns = ['datetime', 'username', 'content'])
except Exception as e:
    print(e)

print("Proses Crawling Data Sudah Selesai")
print("-----")

CPU times: total: 0 ns
Wall time: 0 ns

-----
Mulai Proses Crawling Data
Proses Crawling Data Sudah Selesai
    
```

Gambar 2. Proses Pengumpulan Dataset Menggunakan sncraspe

Pada setiap tahapan *crawling data* pada Twitter hanya mendapatkan akses untuk mengambil dataset sebanyak 100 *tweet* per sekali pengambilan dikarenakan batas limitasi yang ditentukan oleh *library sncraspe*. Oleh karena itu, dilakukan proses *crawling data* berkali-kali hingga mendapatkan jumlah sumber data sesuai kebutuhan. Setelah dataset terkumpul sesuai dengan kebutuhan, dataset tersebut akan digabungkan dan dikonversi menjadi *file* dengan format *\*.CSV (Comma Separated Values)* untuk dilakukan proses *data selection*.

Sumber data yang telah dikumpulkan akan dilakukan proses *data selection* secara manual seperti menghilangkan data yang duplikat, data yang kosong, data yang *error* atau data yang tidak sesuai dengan ketentuan penelitian untuk mencegah terjadinya permasalahan ketika proses *modeling* dan evaluasi. Total dataset yang dikumpul untuk pelaksanaan penelitian ini adalah sebanyak 500 dataset. Dataset yang dihapus karena tidak sesuai dengan ketentuan penelitian dapat dilihat pada Gambar 3.

datetime	username	content	reason
2022-11-02 23:25:21 +00:00	Challuvs	"Rekomendasi tas ransel buat laptop anti air buat kuliah &amp;mp i6" a thread Kamus TV analog sCTV lahodeinbek Gopoke digital gratis ongkir Tokopedia wewoo PURPLE PANIC ON GMA GLOBAL Pempusn Bagan Merah Unsur Majan Chelsea Tidur Zlaco Kebangn. #TheLinkinKt @confirmed https://t.co/AE5lanY01P"	Tweeter ini merupakan promosi, bukan merupakan opini/pesadapat
2022-11-02 13:42:18 +00:00	gracier	Sedih saya tdk koring kt Azista Arena. Time player Malaysia tgl main n per hikan sokongan koring boleh benti tempet dadak semati2 nk nk tgl match Momota. Obvious kt tv lak tu siap pengulas meroboi hadu tu. #PerchedMalaysiaMnien2022	Tweeter ini menggunakan bahasa Melayu, bukan bahasa Indonesia
#NAME?	#NAME?	#NAME?	Tweeter error

Gambar 3. Contoh Dataset yang dihapus

C. Pre-processing

Tahapan pengolahan data (*pre-processing*) adalah langkah mempersiapkan data mentah sebelum akan melakukan proses selanjutnya. Tahapan ini memiliki tujuan untuk dapat menyeragamkan kata dan mengurangi volume kata dari sekumpulan dataset serta memudahkan penelitian karena akan membuat data lebih terstruktur [18]. Proses pengolahan data ini akan melewati 6 (enam) tahapan, yaitu: *Case-folding*, *Punctuation Removal*, *Normalization*, *Stemming*, *Tokenization*, dan *Stop-word Removal*.

D. Data Labelling

Setelah melakukan proses pengolahan data, kemudian dilakukan proses pelabelan data. Pada proses label data menggunakan metode manual dan metode *Lexicon Based* agar dapat membandingkan serta menyelesaikan permasalahan klasifikasi *tweet* agar data dapat masuk ke dalam kelompok yang benar. Pada penelitian ini akan membagi klasifikasi menjadi 2 (dua) kelompok, yaitu pernyataan positif ditandai dengan "1" (satu) dan negatif ditandai dengan "0" (nol).

E. Word Embeddings

Pada proses pembobotan kata akan menggunakan metode *Term Frequency - Inverse Document Frequency* (TF-IDF). Pada dasarnya, TF-IDF bekerja dengan menentukan frekuensi relatif kata-kata dalam suatu hal tertentu dokumen dan dibandingkan dengan proporsi terbalik dari kata itu di seluruh korpus dokumen [19].

$$w_d = f_{w,d} \times \log\left(\frac{|D|}{f_{w,d}}\right) \quad (1)$$

F. Natural Language Processing

Pada implementasi *Natural Language Processing* akan menggunakan *library* WordCloud. WordCloud merupakan salah satu *library* pada pemrograman Python yang digunakan untuk menunjukkan kata yang paling sering digunakan atau sering muncul di kumpulan dataset.

G. Model Algoritma Klasifikasi

- *Multinomial Naive Bayes* adalah pengklasifikasian statistik yang dapat digunakan untuk memprediksi probabilitas keanggotaan suatu kelas [20].

$$P(X_i | H_j) = \frac{\text{count}(X_i, H_j) + 1}{\left(\sum_{x \in V} \text{count}(X, H_j)\right) + |V|} \quad (2)$$

- *Support Vector Machines* adalah algoritma yang

tergolong dalam *Supervised Learning* yang digunakan untuk klasifikasi dan regresi [21].

$$f(X_d) = \sum_{i=1}^{N_{sv}} a_{iyi} (yx_i^T + r)^p + b \quad (3)$$

- *Logistic Regression* adalah metode statistik yang umumnya sering digunakan untuk analisis data dan mendeskripsikan variabel respon dan prediksi [22].

$$f(y_i) = \pi(x_i)^{y_i} (1 - \pi(x_i))^{1-y_i} \quad (4)$$

H. Evaluasi

*Confusion Matrix* adalah metode untuk proses evaluasi menggunakan tabel matriks. Hasil evaluasi akan menghasilkan nilai *accuracy*, *precision*, dan *recall*.

- *Accuracy* adalah tingkat kedekatan data latih dan data uji yang telah diproses [23].

$$\text{accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (5)$$

- *Precision* adalah tingkat ketepatan dari data yang diminta dengan data yang dihasilkan [23].

$$\text{precision} = \frac{TP}{TP + FP} \quad (6)$$

- *Recall* adalah tingkat keberhasilan sebuah sistem dalam klasifikasi [23].

$$\text{recall} = \frac{TP}{TP + FN} \quad (7)$$

- *F1-Score* adalah hasil evaluasi gabungan dari *precision* dan *recall* [23].

$$f_1 - \text{score} = 2 \times \frac{\text{precision} \times \text{recall}}{\text{precision} + \text{recall}} \quad (8)$$

III. HASIL DAN PEMBAHASAN

A. Data Collection

Pada tahapan pengumpulan data, dilakukan proses pengumpulan *dataset* sebagai data penelitian menggunakan *library* *snsrcape*. Tahapan ini melakukan pencarian berdasarkan kata kunci yang berhubungan dengan topik penelitian yaitu "migrasi ke televisi digital". Proses hasil *crawling data* dengan *snsrcape* dapat dilihat pada Gambar 4.



Gambar 4. Proses Pengumpulan Dataset dengan snsrcape

B. Pre-Processing Data

Case-folding adalah tahapan untuk mengkonversikan semua karakter dengan huruf kapital menjadi huruf kecil. Proses case-folding dapat dilihat pada Gambar 5.

Raw Data	Case-folding
Eh udah ASO kan? Ternyata masih bisa nangkap. Apa ternyata TV ny digital?	eh udah aso kan? ternyata masih bisa nangkap. apa ternyata tv ny digital?

Gambar 5. Proses Case-folding

Punctuation Removal adalah proses menghilangkan kata atau angka acak yang tidak dibutuhkan, menghapus karakter dan emoji, spasi berlebih, URL web, kata yang duplikat, serta menghapus tag, retweet, dan hashtag. Proses Punctuation Removal dapat dilihat pada Gambar 6.

Case-folding	Punctuation Removal
mahfid md yakin migrasi tv analog ke digital tak akan timbulkan gejala di masyarakat <a href="https://t.co/dpwp2f2eh">https://t.co/dpwp2f2eh</a> @syamam59 @kemkominfo id ? kalau sudah migrasi ke digital semua kita akan kangen dengan gambar semut .dy .ce9%	mahfid md yakin migrasi tv analog ke digital tak akan timbulkan gejala di masyarakat kalau sudah migrasi ke digital semua kita akan kangen dengan gambar semut

Gambar 6. Proses Punctuation Removal

Normalization adalah proses menormalkan sejumlah kata pada sebuah data menjadi kata-kata yang lebih terstruktur, mengubah bahasa yang tidak baku menjadi baku, memperbaiki kata-kata yang disingkat. Proses Normalization dapat dilihat pada Gambar 7.

Punctuation Removal	Normalization
beli setup box kalau tv kalian ga support tv digital biar kalian tetap bisa nonton tv channel lebih banyak dan gambar lebih jernih	beli setup box kalau televisi kalian tidak mendukung televisi digital biar kalian tetap bisa nonton televisi kanal lebih banyak dan gambar lebih jernih

Gambar 7. Proses Normalization

Stemming adalah proses menghilangkan kata yang berbentuk imbuhan menjadi kata dasar. Proses Stemming dapat dilihat pada Gambar 8.

Normalization	Stemming
mulai rabu keminfo matikan siaran televisi analog secara bertahap	mulai rabu keminfo mati siaran televisi analog secara tahap

Gambar 8. Proses Stemming

Tokenization adalah proses pemotongan kata menjadi satu-satu kata pada setiap data. Setiap potongan kata disebut token yang akan dipisahkan tiap frasa yang dibatasi oleh spasi. Tujuannya agar mempermudah proses pembobotan kata karena pada setiap token akan diberikan nilai berdasarkan korpus yang sudah tersedia. Proses Tokenization dapat dilihat pada Gambar 9.

Stemming	Tokenization
pemerintah resmi henti siaran televisi analog di wilayah jakarta bogor depok tangerang bekasi	{pemerintah, resmi, henti, siaran, televisi, analog, di, wilayah, jakarta, bogor, depok, tangerang, bekasi}

Gambar 9. Proses Tokenization

Stop-word Removal adalah tahapan terakhir dalam pre-processing data. Proses ini menyaring kata penghubung yang

bersifat umum dan tidak berpengaruh kepada proses pembagian kelas sehingga kata penghubung tersebut tidak diperlukan. Proses Stop-word Removal dapat dilihat pada Gambar 10.

Tokenization	Stop-word Removal
{setelah, sempat, undur, dan, oktober, lalu, pemerintahan, lalu, kementerian, kemkominfo, telah, resmi, berhenti, siaran, televisi, analog, atau, analog, switch, off, analog, switch, off, id, wilayah, termasuk, jakarta, bogor, depok, tangerang, bekasi, wib, pulau, wib, lengkap, id}	{undur, oktober, pemerintahan, kementerian, kemkominfo, resmi, berhenti, siaran, televisi, analog, analog, switch, off, analog, switch, off, wilayah, jakarta, bogor, depok, tangerang, bekasi, wib, lengkap}

Gambar 10. Proses Stop-word Removal

C. Data Labelling

Metode Manual

Untuk proses label data secara manual ini bersumber dari penelitian terdahulu yang dilakukan oleh [24][25]. Tahapan ini akan melakukan pelabelan data secara manual yang dilakukan oleh peneliti terhadap setiap dataset yang tersedia, pada penelitian ini hanya akan membagi menjadi dua kelas, yaitu label "0" untuk negatif, dan label "1" untuk positif. Contoh label data secara manual dapat dilihat pada Gambar 11.

Date Time	Username	Cleaned Data	Label
2022-11-02 17:07:51+00:00	aldi_akbar98	terima kasih heuti siaran televisi analog analog switch off televisi digital bersih jernih canggih	1
2022-11-02 23:54:05+00:00	Galihguny_	televiisi analog off bisnis pemerintah biar beli televisi	0

Gambar 11. Proses Label Manual

Metode Lexicon Based

Tahapan ini akan melakukan pelabelan data menggunakan metode Lexicon Based. Karena metode ini hanya tersedia korpus berbahasa Inggris, maka seluruh dataset diterjemahkan menjadi bahasa Inggris terlebih dahulu dengan library googletrans pada Python agar dapat mempermudah proses label data. Kemudian hasil score positif ( $\geq 0$ ) akan diberi label "1" (satu), dan yang negatif ( $< 0$ ) akan diberi label "0" (nol). Contoh label data Lexicon Based pada Gambar 12.

Date Time	Username	Cleaned Data (English)	Score	Label
2022-11-02 17:07:51+00:00	aldi_akbar 98	thank you stop broadcasting analog television analog switch off digital television clean clear sophisticated	0.8481	1
2022-11-02 23:54:57+00:00	Lastri 2108	tv channels are gone buying set top boxes crazy money life is hard in this country people buy set top boxes	-0.0516	0

Gambar 12. Proses Label Lexicon Based

Pada Gambar 13 adalah hasil distribusi pembagian dataset dengan kelas positif dan kelas negatif.

out[5]: 1 391 0 109 Name: manuallabel, dtype: int64	out[25]: 1 399 0 101 Name: sentiment, dtype: int64
---	--

Gambar 13. Distribusi Kelas Label Manual (Kiri) dan Lexicon Based (Kanan)

D. Word Embeddings

Berikut ini adalah hasil output dari pembobotan kata menggunakan metode Term Frequency - Inverse Document Frequency (TF-IDF) yang dapat dilihat pada Gambar 14.



$$precision = \frac{43}{43+6} = \frac{43}{49} = 0,8776$$

$$recall = \frac{43}{43+0} = \frac{43}{43} = 1,00$$

$$f_1 - score = 2 \times \frac{0,8776 \times 1}{0,8776 + 1} = 2 \times \frac{0,8776}{1,8776} = 0,9348$$

Hasil *score confusion matrix* pada pembobotan data latih 80% dan data uji 20%:

$$accuracy = \frac{86+1}{86+1+13+0} = \frac{87}{100} = 0,87$$

$$precision = \frac{86}{86+13} = \frac{86}{99} = 0,8687$$

$$recall = \frac{86}{86+0} = \frac{86}{86} = 1,00$$

$$f_1 - score = 2 \times \frac{0,8687 \times 1}{0,8687 + 1} = 2 \times \frac{0,8687}{1,8687} = 0,9297$$

Hasil *score confusion matrix* pada pembobotan data latih 70% dan data uji 30%:

$$accuracy = \frac{125+1}{125+1+24+0} = \frac{126}{150} = 0,84$$

$$precision = \frac{125}{125+24} = \frac{125}{149} = 0,8389$$

$$recall = \frac{125}{125+0} = \frac{125}{125} = 1,00$$

$$f_1 - score = 2 \times \frac{0,8389 \times 1}{0,8389 + 1} = 2 \times \frac{0,8389}{1,8389} = 0,9124$$

Hasil *score confusion matrix* pada pembobotan data latih 60% dan data uji 40%:

$$accuracy = \frac{166+1}{166+1+33+0} = \frac{167}{200} = 0,835$$

$$precision = \frac{166}{166+33} = \frac{166}{199} = 0,8342$$

$$recall = \frac{166}{166+0} = \frac{166}{166} = 1,00$$

$$f_1 - score = 2 \times \frac{0,8342 \times 1}{0,8342 + 1} = 2 \times \frac{0,8342}{1,8342} = 0,9096$$

Hasil *score confusion matrix* pada pembobotan data latih 50% dan data uji 50%:

$$accuracy = \frac{210+1}{210+1+39+0} = \frac{211}{250} = 0,844$$

$$precision = \frac{210}{210+39} = \frac{210}{249} = 0,8434$$

$$recall = \frac{210}{210+0} = \frac{210}{210} = 1,00$$

$$f_1 - score = 2 \times \frac{0,8434 \times 1}{0,8434 + 1} = 2 \times \frac{0,8434}{1,8434} = 0,915$$

Klasifikasi Algoritma *Multinomial Naïve Bayes* dengan nilai *accuracy* terbaik pada metode manual adalah 80,40%, sedangkan pada metode *Lexicon Based* adalah 88,00%.

### Support Vector Machines

Pada tahap klasifikasi menggunakan *Support Vector Machines* akan dilakukan pengujian sebanyak 5 (lima) kali dengan bobot data *train* dan data *test* yang berbeda agar dapat melihat hasil terbaik dari pengujian dengan ketentuan data *train* dan data *test* dilakukan secara acak. Tabel III merupakan hasil pengujian label manual, sedangkan pada Tabel IV merupakan hasil pengujian *Lexicon Based*.

TABEL III. HASIL KLASIFIKASI SUPPORT VECTOR MACHINES (MANUAL)

Bobot Data		Jumlah Data		Confusion Matrix				Score			
Latih	Uji	Latih	Uji	TP	TN	FP	FN	Accuracy	Precision	Recall	F1-Score
90%	10%	450	50	40	1	9	0	82,00%	81,63%	100,00%	89,89%
80%	20%	400	100	77	2	19	2	79,00%	80,21%	97,47%	88,00%
70%	30%	350	150	114	6	26	4	80,00%	81,43%	96,61%	88,37%
60%	40%	300	200	155	4	36	5	79,50%	81,15%	96,88%	88,32%
50%	50%	250	250	194	6	43	7	80,00%	81,86%	96,52%	88,59%

Hasil *score confusion matrix* pada pembobotan data latih 90% dan data uji 10%:

$$accuracy = \frac{40+1}{40+1+9+0} = \frac{41}{50} = 0,82$$

$$precision = \frac{40}{40+9} = \frac{40}{49} = 0,8163$$

$$recall = \frac{40}{40+0} = \frac{40}{40} = 1,00$$

$$f_1 - score = 2 \times \frac{0,8163 \times 1}{0,8163 + 1} = 2 \times \frac{0,8163}{1,8163} = 0,8989$$

Hasil *score confusion matrix* pada pembobotan data latih 80% dan data uji 20%:

$$accuracy = \frac{77+2}{77+2+19+2} = \frac{79}{100} = 0,79$$

$$precision = \frac{77}{77+19} = \frac{77}{96} = 0,8021$$

$$recall = \frac{77}{77+2} = \frac{77}{79} = 0,9747$$

$$f_1 - score = 2 \times \frac{0,8021 \times 0,9747}{0,8021 + 0,9747} = 2 \times \frac{0,78180687}{1,7768} = 0,88$$

Hasil *score confusion matrix* pada pembobotan data latih 70% dan data uji 30%:

$$accuracy = \frac{114+6}{114+6+26+4} = \frac{120}{150} = 0,80$$

$$precision = \frac{114}{114+26} = \frac{114}{140} = 0,8143$$

$$recall = \frac{114}{114+4} = \frac{114}{118} = 0,9661$$

$$f_1 - score = 2 \times \frac{0,8143 \times 0,9661}{0,8143 + 0,9661} = 2 \times \frac{0,78669523}{1,7804} = 0,8837$$

Hasil *score confusion matrix* pada pembobotan data latih 60% dan data uji 40%:

$$accuracy = \frac{155+4}{155+4+36+5} = \frac{159}{200} = 0,795$$

$$precision = \frac{155}{155+36} = \frac{155}{191} = 0,8115$$

$$recall = \frac{155}{155+5} = \frac{155}{160} = 0,9688$$

$$f_1 - score = 2 \times \frac{0,8115 \times 0,9688}{0,8115 + 0,9688} = 2 \times \frac{0,7861812}{1,7803} = 0,8832$$

Hasil *score confusion matrix* pada pembobotan data latih 50% dan data uji 50%:

$$accuracy = \frac{194+6}{194+6+43+7} = \frac{200}{250} = 0,80$$

$$precision = \frac{194}{194+43} = \frac{194}{237} = 0,8186$$

$$recall = \frac{194}{194+7} = \frac{194}{201} = 0,9652$$

$$f_1 - score = 2 \times \frac{0,8186 \times 0,9652}{0,8186 + 0,9652} = 2 \times \frac{0,79011272}{1,7838} = 0,8859$$

TABEL IV. HASIL KLASIFIKASI SUPPORT VECTOR MACHINES (LEXICON BASED)

Bobot Data		Jumlah Data		Confusion Matrix				Score			
Latih	Uji	Latih	Uji	TP	TN	FP	FN	Accuracy	Precision	Recall	F1-Score
90%	10%	450	50	43	4	3	0	94,00%	93,48%	100,00%	96,63%
80%	20%	400	100	86	5	9	0	91,00%	90,53%	100,00%	95,03%
70%	30%	350	150	121	9	16	4	86,67%	88,32%	96,80%	92,37%
60%	40%	300	200	158	12	22	8	85,00%	87,78%	95,18%	91,33%
50%	50%	250	250	201	14	26	9	86,00%	88,55%	95,71%	91,99%

Hasil *score confusion matrix* pada pembobotan data latih 90% dan data uji 10%:

$$accuracy = \frac{43+4}{43+4+3+0} = \frac{47}{50} = 0,94$$

$$precision = \frac{43}{43+3} = \frac{43}{46} = 0,9348$$

$$recall = \frac{43}{43+0} = \frac{43}{43} = 1,00$$

$$f_1 - score = 2 \times \frac{0,9348 \times 1}{0,9348 + 1} = 2 \times \frac{0,9348}{1,9348} = 0,9663$$

Hasil *score confusion matrix* pada pembobotan data latih 80% dan data uji 20%:

$$accuracy = \frac{86+5}{86+5+9+0} = \frac{91}{100} = 0,91$$

$$precision = \frac{86}{86+9} = \frac{86}{95} = 0,9053$$

$$recall = \frac{86}{86+0} = \frac{86}{86} = 1,00$$

$$f_1 - score = 2 \times \frac{0,9053 \times 1}{0,9053 + 1} = 2 \times \frac{0,9053}{1,9053} = 0,9503$$

Hasil *score confusion matrix* pada pembobotan data latih 70% dan data uji 30%:

$$accuracy = \frac{121+9}{121+9+16+4} = \frac{130}{150} = 0,8667$$

$$precision = \frac{121}{121+16} = \frac{121}{137} = 0,8832$$

$$recall = \frac{121}{121+4} = \frac{121}{125} = 0,968$$

$$f_1 - score = 2 \times \frac{0,8832 \times 0,968}{0,8832 + 0,968} = 2 \times \frac{0,8549376}{1,8512} = 0,9237$$

Hasil *score confusion matrix* pada pembobotan data latih 60% dan data uji 40%:

$$accuracy = \frac{158+12}{158+12+22+8} = \frac{170}{200} = 0,85$$

$$precision = \frac{158}{158+22} = \frac{158}{180} = 0,8778$$

$$recall = \frac{158}{158+8} = \frac{158}{166} = 0,9518$$

$$f_1 - score = 2 \times \frac{0,8778 \times 0,9518}{0,8778 + 0,9518} = 2 \times \frac{0,83549004}{1,8296} = 0,9133$$

Hasil *score confusion matrix* pada pembobotan data latih 50% dan data uji 50%:

$$accuracy = \frac{201+14}{201+14+26+9} = \frac{215}{250} = 0,86$$

$$precision = \frac{201}{201+16} = \frac{201}{217} = 0,8855$$

$$recall = \frac{201}{201+16} = \frac{201}{217} = 0,9571$$

$$f_1 - score = 2 \times \frac{0,8855 \times 0,9571}{0,8855 + 0,9571} = 2 \times \frac{0,84751205}{1,8426} = 0,9199$$

Klasifikasi Algoritma *Support Vector Machines* dengan nilai *accuracy* terbaik pada metode manual adalah 82,00%, sedangkan pada metode *Lexicon Based* adalah 94,00%.

#### Logistic Regression

Pada tahap klasifikasi menggunakan *Logistic Regression* akan dilakukan pengujian sebanyak 5 (lima) kali dengan bobot data *train* dan data *test* yang berbeda agar dapat melihat hasil terbaik dari pengujian dengan ketentuan data *train* dan data *test* dilakukan secara acak. Tabel V merupakan hasil pengujian label manual, sedangkan pada Tabel VI merupakan hasil pengujian *Lexicon Based*.

TABEL V. HASIL KLASIFIKASI LOGISTIC REGRESSION (MANUAL)

Bobot Data		Jumlah Data		Confusion Matrix				Score			
Latih	Uji	Latih	Uji	TP	TN	FP	FN	Accuracy	Precision	Recall	F1-Score
90%	10%	450	50	40	0	10	0	80,00%	80,00%	100,00%	88,89%
80%	20%	400	100	79	1	20	0	80,00%	79,80%	100,00%	88,77%
70%	30%	350	150	118	1	31	0	79,33%	79,19%	100,00%	88,39%
60%	40%	300	200	160	1	39	0	80,50%	80,40%	100,00%	89,14%
50%	50%	250	250	201	1	48	0	80,80%	80,72%	100,00%	89,33%

Hasil *score confusion matrix* pada pembobotan data latih 90% dan data uji 10%:

$$accuracy = \frac{40+0}{40+0+10+0} = \frac{40}{50} = 0,80$$

$$precision = \frac{40}{40+10} = \frac{40}{50} = 0,80$$

$$recall = \frac{40}{40+0} = \frac{40}{40} = 1,00$$

$$f_1 - score = 2 \times \frac{0,8 \times 1}{0,8 + 1} = 2 \times \frac{0,8}{1,8} = 0,8889$$

Hasil *score confusion matrix* pada pembobotan data latih 80% dan data uji 20%:

$$accuracy = \frac{79+1}{79+1+20+0} = \frac{80}{100} = 0,80$$

$$precision = \frac{79}{79+20} = \frac{79}{99} = 0,798$$

$$recall = \frac{79}{79+0} = \frac{79}{79} = 1,00$$

$$f_1 - score = 2 \times \frac{0,798 \times 1}{0,798+1} = 2 \times \frac{0,798}{1,798} = 0,8877$$

Hasil *score confusion matrix* pada pembobotan data latih 70% dan data uji 30%:

$$accuracy = \frac{118+1}{118+1+31+0} = \frac{119}{150} = 0,7933$$

$$precision = \frac{118}{118+31} = \frac{118}{149} = 0,7919$$

$$recall = \frac{118}{118+0} = \frac{118}{118} = 1,00$$

$$f_1 - score = 2 \times \frac{0,7919 \times 1}{0,7919+1} = 2 \times \frac{0,7919}{1,7919} = 0,8839$$

Hasil *score confusion matrix* pada pembobotan data latih 60% dan data uji 40%:

$$accuracy = \frac{160+1}{160+1+39+0} = \frac{161}{200} = 0,805$$

$$precision = \frac{160}{160+39} = \frac{160}{199} = 0,804$$

$$recall = \frac{160}{160+0} = \frac{160}{160} = 1,00$$

$$f_1 - score = 2 \times \frac{0,804 \times 1}{0,804+1} = 2 \times \frac{0,804}{1,804} = 0,8914$$

Hasil *score confusion matrix* pada pembobotan data latih 50% dan data uji 50%:

$$accuracy = \frac{201+1}{201+1+48+0} = \frac{202}{250} = 0,808$$

$$precision = \frac{201}{201+48} = \frac{201}{249} = 0,8072$$

$$recall = \frac{201}{201+0} = \frac{201}{201} = 1,00$$

$$f_1 - score = 2 \times \frac{0,8072 \times 1}{0,8072+1} = 2 \times \frac{0,8072}{1,8072} = 0,8933$$

$$accuracy = \frac{43+2}{43+2+5+0} = \frac{45}{50} = 0,90$$

$$precision = \frac{43}{43+5} = \frac{43}{48} = 0,8958$$

$$recall = \frac{43}{43+0} = \frac{43}{43} = 1,00$$

$$f_1 - score = 2 \times \frac{0,8958 \times 1}{0,8958+1} = 2 \times \frac{0,8958}{1,8958} = 0,9451$$

Hasil *score confusion matrix* pada pembobotan data latih 80% dan data uji 20%:

$$accuracy = \frac{86+4}{86+4+10+0} = \frac{90}{100} = 0,90$$

$$precision = \frac{86}{86+10} = \frac{86}{96} = 0,8957$$

$$recall = \frac{86}{86+0} = \frac{86}{86} = 1,00$$

$$f_1 - score = 2 \times \frac{0,8957 \times 1}{0,8957+1} = 2 \times \frac{0,8957}{1,8957} = 0,945$$

Hasil *score confusion matrix* pada pembobotan data latih 70% dan data uji 30%:

$$accuracy = \frac{123+5}{123+5+20+2} = \frac{128}{150} = 0,8533$$

$$precision = \frac{123}{123+20} = \frac{123}{143} = 0,8601$$

$$recall = \frac{123}{123+2} = \frac{123}{125} = 0,984$$

$$f_1 - score = 2 \times \frac{0,8601 \times 0,984}{0,8601+0,984} = 2 \times \frac{0,8463384}{1,8441} = 0,9179$$

Hasil *score confusion matrix* pada pembobotan data latih 60% dan data uji 40%:

$$accuracy = \frac{163+7}{163+7+27+3} = \frac{170}{200} = 0,85$$

$$precision = \frac{163}{163+27} = \frac{163}{190} = 0,8579$$

$$recall = \frac{163}{163+3} = \frac{163}{166} = 0,9819$$

$$f_1 - score = 2 \times \frac{0,8579 \times 0,9819}{0,8579+0,9819} = 2 \times \frac{0,84237201}{1,8398} = 0,9157$$

Hasil *score confusion matrix* pada pembobotan data latih 50% dan data uji 50%:

$$accuracy = \frac{206+7}{206+7+33+4} = \frac{213}{250} = 0,852$$

$$precision = \frac{206}{206+33} = \frac{206}{239} = 0,8619$$

$$recall = \frac{206}{206+4} = \frac{206}{210} = 0,981$$

$$f_1 - score = 2 \times \frac{0,8619 \times 0,981}{0,8619+0,981} = 2 \times \frac{0,8455239}{1,8429} = 0,9176$$

Klasifikasi Algoritma *Logistic Regression* dengan nilai *accuracy* terbaik pada metode manual adalah 80,80%,

TABEL VI. HASIL KLASIFIKASI LOGISTIC REGRESSION (LEXICON BASED)

Bobot Data		Jumlah Data		Confusion Matrix				Score			
Latih	Uji	Latih	Uji	TP	TN	FP	FN	Accuracy	Precision	Recall	F1-Score
90%	10%	450	50	43	2	5	0	90,00%	89,58%	100,00%	94,51%
80%	20%	400	100	86	4	10	0	90,00%	89,57%	100,00%	94,50%
70%	30%	350	150	123	5	20	2	85,33%	86,01%	98,40%	91,79%
60%	40%	300	200	163	7	27	3	85,00%	85,79%	98,19%	91,57%
50%	50%	250	250	206	7	33	4	85,20%	86,19%	98,10%	91,76%

Hasil *score confusion matrix* pada pembobotan data latih 90% dan data uji 10%:



sedangkan pada metode *Lexicon Based* adalah 90,00%.

#### IV. KESIMPULAN

Pada penelitian ini, *dataset* yang diambil dari media sosial Twitter sejumlah 500 data mengenai migrasi televisi digital dan dilakukan perbandingan tiga algoritma klasifikasi, yaitu *Multinomial Naïve Bayes*, *Support Vector Machines*, dan *Logistic Regression*. Hasil *accuracy* terbaik yang dihasilkan oleh *Multinomial Naïve Bayes* adalah 88,00%, *Support Vector Machines* adalah 94,00%, dan *Logistic Regression* adalah 90,00%. Hasil *precision* terbaik yang dihasilkan oleh *Multinomial Naïve Bayes* adalah 87,76%, *Support Vector Machines* adalah 93,48%, dan *Logistic Regression* adalah 89,58%. Hasil *recall* terbaik adalah dengan nilai 100,00% untuk semua jenis algoritma klasifikasi yang menunjukkan bahwa sistem berhasil mengklasifikasi semua dataset dengan sangat baik, kemudian hasil *f1-score* terbaik yang dihasilkan oleh *Multinomial Naïve Bayes* 93,48%, *Support Vector Machines* adalah 96,63%, dan *Logistic Regression* adalah 94,51%.

Hasil nilai *confusion matrix* terbaik yang dijelaskan sebelumnya adalah hasil pengujian dengan metode *Lexicon Based*, dan dapat disimpulkan bahwa penggunaan metode manual tidak memiliki hasil yang lebih baik dibandingkan dengan metode *Lexicon Based* untuk seluruh algoritma klasifikasi. Algoritma *Support Vector Machines* memiliki hasil *confusion matrix* (*accuracy*, *precision*, dan *f1-score*) yang lebih baik dibandingkan dengan *Multinomial Naïve Bayes* dan *Logistic Regression* karena terlihat dari nilai TP dan TN yang lebih besar dan FP dan FN yang lebih kecil dibandingkan dengan *Multinomial Naïve Bayes* dan *Logistic Regression* walaupun tidak terlalu jauh jumlah nilai perbedaannya.

Berdasarkan hasil penelitian yang sudah disimpulkan, diharapkan hasil penelitian ini dapat menjadi sebuah referensi dan dapat dilakukan pengembangan lagi untuk kedepannya, seperti menggunakan perbandingan algoritma klasifikasi yang lain atau menggunakan metode yang berbeda untuk pengujian analisis sentimen dengan harapan mendapatkan hasil *accuracy*, *precision*, *recall*, dan *f1-score* yang semakin baik, serta sebagai saran juga dapat menambahkan golongan kelas yang lebih detail lagi agar data analisis sentimen semakin akurat, atau juga dapat dikembangkan menjadi sebuah sistem aplikasi berbasis web ataupun *mobile* untuk klasifikasi analisis sentimen.

#### DAFTAR PUSTAKA

- [1] M. Oktaviana, Z. A. Achmad, H. Arviani, and K. Kusnarto, "Budaya Komunikasi Virtual di Twitter dan Tiktok: Perluasan Makna Kata Estetik," *Satwika: Kajian Ilmu Budaya dan Perubahan Sosial*, vol. 5, no. 2, pp. 173-186. 2021.
- [2] M. Alaydrus, "Digital Dividend pada Migrasi TV Analog ke TV Digital-Prospek dan Dilema," . *InComTech: Jurnal Telekomunikasi dan Komputer*, vol. 1, no. 1, pp. 1-9. 2009.
- [3] M. Firdaus, "Analisis Inovasi Industri Televisi Menghadapi Migrasi Televisi Digital (Studi Kasus Televisi Lokal Di Provinsi Bengkulu)," *In Conference on Economic and Business Innovation (CEBI)*, pp. 783-800. 2021.
- [4] S. Budhirianto, "Sikap Masyarakat terhadap Kampanye Televisi Digital pada Media Televisi," *Jurnal Penelitian Komunikasi Dan Opini Publik*, vol. 18, no. 3. 2014.
- [5] V. Zuliana, G. Garno, and I. Maulana, "Analisis Sentimen Program Migrasi TV Digital Menggunakan Algoritma Naive Bayes dengan Chi Square," *Jurnal Informasi dan Komputer*, vol. 10, no. 2, pp. 90-95. 2022.
- [6] S. Tiwari, A. Verma., P. Garg., and D. Bansal, "Social Media Sentiment Analysis on Twitter Datasets," *In 2020 6th International Conference on Advanced Computing and Communication Systems (ICACCS)*, pp. 925-927. 2020.
- [7] A. Shelar, and C. Y. Huang, "Sentiment Analysis of Twitter Data," *In 2018 International Conference on Computational Science and Computational Intelligence (CSCI)*, pp. 1301-1302. 2018.
- [8] C. Kariya, and P. Khodke, "Twitter Sentiment Analysis," *In 2020 International Conference for Emerging Technology (INCET)*, pp. 1-3. 2020.
- [9] L. Pasteur, and R. Koch, "Mining Twitter Data on COVID-19 for Sentiment Analysis and Frequent Patterns Discovery Habiba," *Algiers Univ. vol. 74*, no. 1934, pp. 1-3, 2020.
- [10] S. Tam, R. B. Said., and Ö. Ö. Tanrıöver, "A ConvBiLSTM Deep Learning Model-Based Approach for Twitter Sentiment Classification," pp. 41283-41293. 2021.
- [11] F. F. Rachman, and S. Pramana, "Analisis Sentimen Pro dan Kontra Masyarakat Indonesia Tentang Vaksin COVID-19 pada Media Sosial Twitter," *Indonesian of Health Information Management Journal*, vol. 8, no. 2, pp. 100-109. 2020.
- [12] Yuyun, N. Hidayah, and S. Sahibu, "Algoritma Multinomial Naïve Bayes Untuk Klasifikasi Sentimen Pemerintah Terhadap Penanganan COVID-19 Menggunakan Data Twitter," *Jurnal Resti (Rekayasa Sistem dan Teknologi Informasi)*, vol. 5, no. 4, pp. 820-826. 2021.
- [13] N. A. Susanti, M. Walid, and H. Hoiriyah, "Klasifikasi Data Tweet Ujaran Kebencian Di Media Sosial Menggunakan Naive Bayes Classifier", *JATI (Jurnal Mahasiswa Teknik Informatika)*, vol. 6, no. 2, pp. 538-543. 2022.
- [14] P. Arsi, and R. Waluyo, "Analisis Sentimen Wacana Pemindahan Ibu Kota Indonesia Menggunakan Algoritma Support Vector Machine (SVM)," *Jurnal Teknologi Informasi dan Ilmu Komputer*, vol. 8, no. 1, pp. 147. 2021.
- [15] D. A. Ramadhan, and E. B. Setiawan, "Analisis Sentimen Program Acara di SCTV pada Twitter Menggunakan Metode Naive Bayes dan Support Vector Machine," *eProceedings of Engineering*, vol. 6, no. 2. 2019.
- [16] A. K. Santoso, A. Noviriandini, A. Kurniasih, B. Wicaksono, and A. Nuryanto, "Klasifikasi Persepsi Pengguna Twitter Terhadap Kasus COVID-19 menggunakan Metode Logistic Regression," *Jurnal Informatika Kaputama (JIK)*, vol. 5, no. 2, pp. 234-241. 2021.
- [17] A. Novantika, and S. Sugiman, "Analisis Sentimen Ulasan Pengguna Aplikasi Video Conference Google Meet menggunakan Metode SVM dan Logistic Regression," *In PRISMA, Prosiding Seminar Nasional Matematika*, vol. 5, pp. 808-813. 2022.
- [18] T. Wulandari, *Klasifikasi Jenis Emosi dari Tweet Berbahasa Indonesia Menggunakan Metode Support Vector Machine*. Doctoral Dissertation, Universitas Islam Negeri Sultan Syarif Kasim: Riau, 2018.
- [19] M. W. Putri, A. Muchayan, and M. Kamisutara, "Sistem Rekomendasi Produk Pena Eksklusif Menggunakan Content-Based Filtering dan TF-IDF," *Journal of Information Technology and Computer Science*, vol. 5, no. 3, pp. 229-236. 2020.
- [20] F. Ariani, and A. Taufik, "Perbandingan Metode Klasifikasi Data Mining untuk Prediksi Tingkat Kepuasan Pelanggan Telkomsel Prabayar," *SATIN-Sains dan Teknologi Informasi*, vol. 6, no. 2, pp. 46-55. 2020.
- [21] K. Kelvin, J. Banjarnahor, M. N. Nababan, S. H. Sinurat, "Analisis Perbandingan Sentimen Corona Virus Disease-2019 (COVID19) pada Twitter Menggunakan Metode Logistic Regression dan Support Vector Machine," *Jurnal Sistem Informasi dan Ilmu Komputer Prima*, vol. 5, no. 2, pp. 47-52. 2022.
- [22] A. Bimantara, and T. A. Dina, "Klasifikasi Web Berbahaya Menggunakan Metode Logistic Regression," *In Annual Research Seminar (ARS)*, vol. 4, no. 1, pp. 173-177. 2019.
- [23] A. R. Rizkirobby, M. Nasrun, and R. A. Nugrahaeni, "Deteksi Ujaran Ancaman Berbasis Website pada Postingan Media Sosial Twitter Menggunakan Metode Naïve Bayes," *eProceedings of Engineering*, vol. 8, no. 1. 2021.
- [24] W. Willianto, I. A. Musdar, J. Junaedy, and H. Angriani, "Implementasi Teori Naive Bayes dalam Klasifikasi Ujaran Kebencian di Facebook," *Jurnal Informatika Universitas Pamulang*, vol. 6, no. 4, pp. 666-671. 2021.
- [25] A. S. Maulana, *Klasifikasi Ujaran Kebencian dan Bahasa Kasar di Twitter Menggunakan Metode Attention Based Recurrent Neural Network*. Doctoral Dissertation UPN "Veteran Yogyakarta". Yogyakarta: 2022.