

BAB II

TINJAUAN PUSTAKA

2.1. Analisis Sentimen

Analisis sentimen adalah bidang studi yang menganalisis opini, emosi, dan sikap orang dalam bentuk teks (Mittal and Patidar, 2019). Tugas dasar dalam analisis sentimen adalah mengklasifikasikan teks dalam sebuah kalimat atau dokumen dan kemudian menentukan apakah pendapat yang diungkapkan dalam kalimat atau dokumen tersebut positif, negatif atau netral. Analisis sentimen juga dapat mengungkapkan perasaan emosional sedih, senang, atau marah. Ekspresi atau sentimen mengacu pada fokus topik tertentu, pernyataan pada satu topik mungkin memiliki arti yang berbeda untuk pernyataan yang sama pada subjek yang berbeda.

2.2. Twitter

Twitter adalah situs web yang dimiliki dan dioperasikan oleh *Twitter Inc.*, yang menawarkan jejaring sosial berupa mikro blog yang memungkinkan penggunanya untuk mengirim dan membaca pesan twit (Leslie *et al.*, 2017). Twit adalah teks tertulis hingga 280 karakter yang ditampilkan di halaman profil pengguna (Boot *et al.*, 2019). Twit dapat dilihat oleh publik, tetapi pengirim dapat membatasi pengiriman pesan ke daftar teman saja. Pengguna dapat melihat twit pengguna lain yang dikenal sebagai *follower*/pengikut.

2.3. Machine Learning

Pembelajaran mesin (*Machine Learning/ML*) adalah cabang dari kecerdasan buatan (*Artificial Intelligence/AI*) dan ilmu komputer yang berfokus pada

penggunaan data dan algoritma untuk meniru cara manusia belajar dan secara bertahap meningkatkan akurasi (IBM, 2020). Dengan kata lain, pembelajaran mesin adalah studi ilmiah tentang algoritma yang digunakan sistem komputer untuk melakukan tugas tertentu tanpa diprogram secara eksplisit.

2.3.1. Naïve Bayes

Algoritma *Naïve Bayes* (NB) merupakan algoritma yang digunakan untuk mencari nilai probabilitas tertinggi untuk mengklasifikasikan data uji pada kategori yang paling sesuai (Aljameel *et al.*, 2021). Metode *Naïve Bayes* merupakan salah satu metode yang dapat mengklasifikasikan teks.

$$P(A|B) = \frac{P(B|A).P(A)}{P(B)} \quad (2.1)$$

dengan keterangan sebagai berikut:

A = hipotesis data A yang merupakan kelas yang spesifik,

B = data dengan kelas yang belum diketahui,

$P(A|B)$ = probabilitas hipotesis A berdasarkan kondisi B,

$P(B|A)$ = probabilitas B dalam hipotesis A.

$P(A)$ = probabilitas hipotesis A,

$P(B)$ = probabilitas B.

Kelebihan *Naïve Bayes* adalah algoritmanya sederhana tetapi memiliki akurasi yang tinggi (Septiani, 2017). Ada dua tahap dalam klasifikasi tweet.

Tahap pertama adalah pelatihan pada twit yang sudah diketahui kategorinya. Sedangkan tahap kedua adalah proses pengklasifikasian twit yang belum diketahui kategorinya.

Tabel 2.1 Probabilitas Kosakata Positif

Kosakata	f	P(Positif)
menteri	2	0.056
pupr	2	0.056
bangun	2	0.056
infrastruktur	2	0.056
bidang	2	0.056
sumber	2	0.056
daya	2	0.056
air	2	0.056
dukung	2	0.056
kawasan	2	0.056
mukim	2	0.056
rumah	2	0.056
mantap	2	0.056
upaya	2	0.056
bagus	2	0.056
banjir	1	0.028
air	1	0.028
lumpuh	1	0.028
ekonomi	1	0.028
ganggu	1	0.028
sanitasi	1	0.028

Tabel 2.2 Probabilitas Kosakata Negatif

Kosakata	f	P(Negatif)
banjir	2	0.074
air	2	0.074
lumpuh	2	0.074
ekonomi	2	0.074
ganggu	2	0.074
sanitasi	2	0.074
menteri	1	0.037

pupr	1	0.037
bangun	1	0.037
infrastruktur	1	0.037
bidang	1	0.037
sumber	1	0.037
daya	1	0.037
air	1	0.037
dukung	1	0.037
kawasan	1	0.037
mukim	1	0.037
rumah	1	0.037
mantap	1	0.037
upaya	1	0.037
bagus	1	0.037

Query Uji: “menteri dukung bangun sanitasi.”

Dengan menggunakan rumus probabilitas maka didapatkan perhitungan sebagai berikut.

$$P(Q|Positif) = 0,5 \times 0,056 \times 0,056 \times 0,056 \times 0,028 = 2,38 \times 10^{-6}$$

$$P(Q|Negatif) = 0,5 \times 0,037 \times 0,037 \times 0,037 \times 0,074 = 1,88 \times 10^{-6}$$

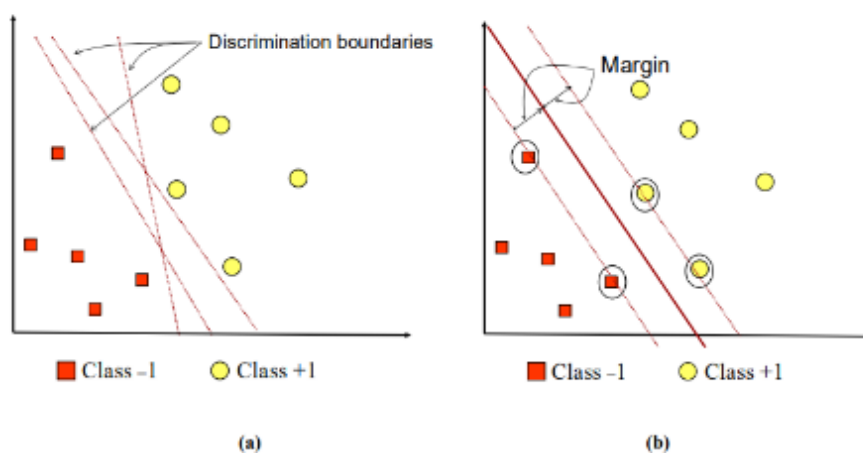
Berdasarkan perhitungan di atas maka probabilitas query uji diprediksi berlabel positif lebih besar daripada probabilitas query uji diprediksi berlabel negatif. Maka berdasarkan perhitungan tersebut query uji diprediksi berlabel positif.

2.3.2. *Support Vector Machine*

Support Vector Machine (SVM) adalah suatu teknik untuk melakukan sebuah prediksi, baik klasifikasi maupun regresi yang terinspirasi dari teori

pembelajaran statistik. SVM dibangun dengan model yang cukup kompleks namun cukup sederhana untuk dianalisis secara matematis.

Ide dasar dari pembelajaran SVM adalah untuk memecahkan sebuah *hyperplane* pemisah yang dapat membagi kumpulan data pelatihan dengan benar dan memiliki interval geometris terbesar (Wang and Zhao, 2020).



Gambar 2.1 SVM berusaha menemukan *hyperplane* terbaik

Sumber: (Nugroho, Witarto and Handoko, 2003)

Hyperplane berguna untuk memisahkan dua kelompok kelas yang masing-masing kelas memiliki pola tersendiri (Luqyana, 2018). Dalam hal ini, SVM akan melakukan proses klasifikasi menggunakan *hyperplane* yang terdapat pada data seperti yang ditunjukkan pada Gambar 2.1.

Data dinotasikan sebagai $\vec{x}_l \in R^2$ sedangkan label masing-masing dinotasikan $y_i \in \{-1, +1\}$ untuk $i = 1, 2, \dots, l$ yang mana l adalah banyaknya data. Asumsi kedua kelas -1 dan +1 dapat terpisah secara sempurna oleh *hyperplane* berdimensi d yang didefinisikan:

$$\vec{w} \cdot \vec{x} + b = 0 \quad (2.2)$$

Pola \vec{x}_l yang termasuk kelas -1 (sampel negatif) dapat dirumuskan sebagai pola yang memenuhi pertidaksamaan

$$\vec{w} \cdot \vec{x} + b \leq -1 \quad (2.3)$$

Sedangkan pola \vec{x}_l yang termasuk kelas +1 (sampel positif) memenuhi pertidaksamaan

$$\vec{w} \cdot \vec{x} + b \geq +1 \quad (2.4)$$

Dengan keterangan sebagai berikut:

\vec{w} = vector pembobot

b = *vector bias*

pemaksimalan jarak terdekat antara hyperplane dengan pattern dilakukan dengan menghitung margin yang mana dirumuskan dengan $\frac{1}{\|\vec{w}\|}$.

2.3.3. *K-Nearest Neighbour*

K-Nearest Neighbor (KNN) merupakan salah satu metode machine learning yang mengklasifikasikan objek berdasarkan data pembelajaran yang paling dekat dengan objek tersebut (Wisnu, Afif and Ruldevyani, 2020). *K-Nearest Neighbor* adalah metode untuk mengklasifikasikan objek baru berdasarkan (K) tetangga terdekat. mendukung pengambilan keputusan, akan terlihat bahwa mayoritas keputusan teman atau tetangga. Tetangga bisa dipilih berdasarkan kedekatannya. Dekat atau jauhnya tetangga biasanya dihitung berdasarkan Euclidean Distance, atau bisa juga menggunakan rumus jarak yang lain. Dimana hasil sampel uji baru diklasifikasikan berdasarkan mayoritas kategori

K-Nearest Neighbor. Jadi K-Nearest Neighbor dari sebuah instance x didefinisikan sebagai instance K yang memiliki jarak terkecil (kedekatan terbesar, *nearest*) ke x . Dekat atau jauhnya tetangga biasanya dihitung berdasarkan Euclidean Distance yang direpresentasikan dalam persamaan 1 sebagai berikut (Kartika et al., 2017)

$$D(a, b) = \sqrt{\sum_{k=1}^d (a_k - b_k)^2} \quad (2.5)$$

Dengan keterangan sebagai berikut:

$D(a, b)$ = jarak Euclidean

d = dimensi data variable bebas

a_k = data testing ke- k

b_k = data training ke- k

Tujuan dari algoritma ini mengklasifikasikan objek baru, atribut dan pelatihan sampel Nilai k terbaik untuk algoritma ini tergantung pada data. Pada dasarnya, nilai k yang tinggi akan mengurangi efek noise pada klasifikasi, tetapi membuat batas antara setiap klasifikasi semakin kabur. (Wisnu, Afif and Ruldevyani, 2020)

Tabel 2.3 Contoh dokumen text

Dokumen (n)	Text	Label
1	menteri pupr bangun infrastruktur bidang sumber daya air dukung kawasan mukim rumah mantap upaya bagus	Positif
2	banjir air lumpuh ekonomi ganggu sanitasi	Negatif

Query Uji: “menteri bumh erick thohir nyata komitmen dukung program bijak merdeka ajar kampus merdeka mkbm”

Tabel 2.4 Hasil Pembobotan

Kosakata	Tf(D1)	Tf(D2)	Tf(Q)	Df	D/df	idf	W(D1)	W(D2)	W(Q)
menteri	1	0	1	2	1.5	0.176	0.176	0.000	0.176
pupr	1	0	0	1	3	0.477	0.477	0.000	0.000
bangun	1	0	0	1	3	0.477	0.477	0.000	0.000
infrastruktur	1	0	0	1	3	0.477	0.477	0.000	0.000
bidang	1	0	0	1	3	0.477	0.477	0.000	0.000
sumber	1	0	0	1	3	0.477	0.477	0.000	0.000
daya	1	0	0	1	3	0.477	0.477	0.000	0.000
air	1	1	0	2	1.5	0.176	0.176	0.176	0.000
dukung	1	0	1	2	1.5	0.176	0.176	0.000	0.176
kawasan	1	0	0	1	3	0.477	0.477	0.000	0.000
mukim	1	0	0	1	3	0.477	0.477	0.000	0.000
rumah	1	0	0	1	3	0.477	0.477	0.000	0.000
mantap	1	0	0	1	3	0.477	0.477	0.000	0.000
upaya	1	0	0	1	3	0.477	0.477	0.000	0.000
bagus	1	0	0	1	3	0.477	0.477	0.000	0.000
banjir	0	1	0	1	3	0.477	0.000	0.477	0.000
lumpuh	0	1	0	1	3	0.477	0.000	0.477	0.000
ekonomi	0	1	0	1	3	0.477	0.000	0.477	0.000
ganggu	0	1	0	1	3	0.477	0.000	0.477	0.000
sanitasi	0	1	0	1	3	0.477	0.000	0.477	0.000
bumh	0	0	1	1	3	0.477	0.000	0.000	0.477
erick	0	0	1	1	3	0.477	0.000	0.000	0.477
thohir	0	0	1	1	3	0.477	0.000	0.000	0.477
nyata	0	0	1	1	3	0.477	0.000	0.000	0.477
komitmen	0	0	1	1	3	0.477	0.000	0.000	0.477
program	0	0	1	1	3	0.477	0.000	0.000	0.477
bijak	0	0	1	1	3	0.477	0.000	0.000	0.477
merdeka	0	0	1	1	3	0.477	0.000	0.000	0.477
ajar	0	0	1	1	3	0.477	0.000	0.000	0.477
kampus	0	0	2	1	3	0.477	0.000	0.000	0.954

mbkm	0	0	1	1	3	0,477	0,000	0,000	0,477
Total							6,254	2,562	6,078

Dengan menggunakan rumus 2.5 maka didapatkan jarak untuk setiap dokumen dibandingkan dengan query yang diuji.

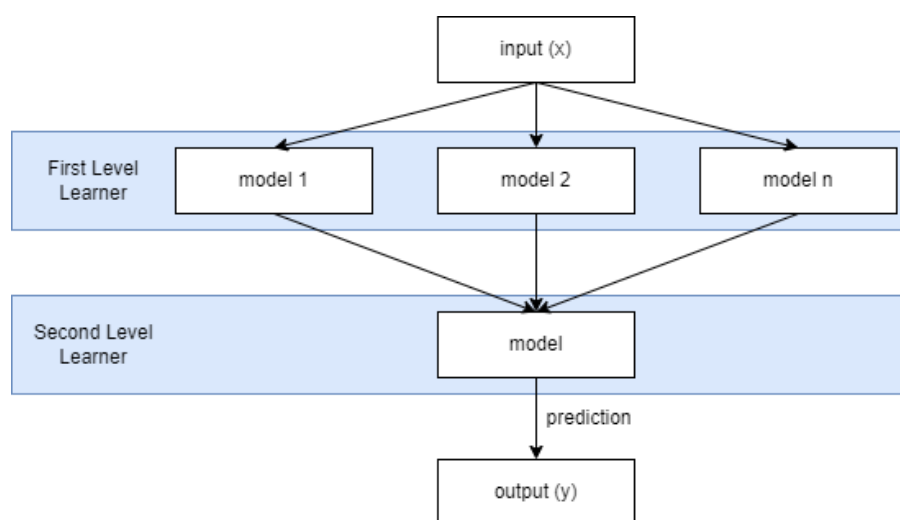
$$d(Q, D1) = \sqrt{(6,078 - 6,254)^2} = 0,176$$

$$d(Q, D2) = \sqrt{(6,078 - 2,562)^2} = 3,516$$

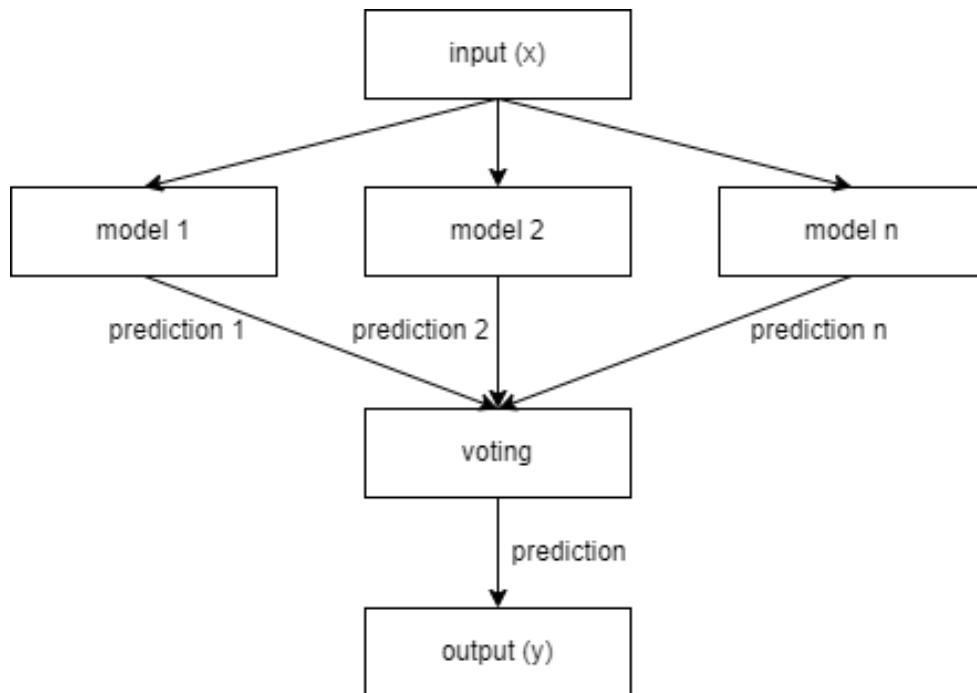
maka didapatkan jarak Q ke D1 bernilai 0,176 dan jarak Q ke D2 bernilai 3,516. Berdasarkan jarak terdekat, maka dari algoritma ini akan menghasilkan prediksi bahwa Q merupakan berlabel Positif.

2.3.4. Ensemble Machine Learning Classifier

Ensemble Machine Learning Classifier adalah model komposit, menggabungkan prediksi dari beberapa model lain (Saleena, 2018). *Ensemble* dikenal lebih akurat daripada model tunggal. Pada Penelitian ini akan membandingkan akurasi dari *ensemble machine learning classifier* tipe *stacking* (Gambar 2.2) dan *voting* (Gambar 2.3).



Gambar 2.2. Ensemble Machine Learning Classifier - Stacking



Gambar 2.3. Ensemble Machine Learning Classifier - Voting

Ensemble stacking classifier yang ditunjukkan oleh Gambar 2.2 memiliki 2 level pembelajar. Pada level 1, terdapat beberapa model algoritma tunggal yang melakukan pemodelan pada tahap pertama. Dilanjutkan kepada level 2 yang hanya memiliki sebuah algoritma tunggal sebagai *meta learner* yang bertugas menggabungkan hasil dari level 1. Setelah data melewati level 2 barulah proses pemodelan ini mendapatkan akurasi prediksi. Selanjutnya adalah *ensemble voting classifier* yang ditunjukkan oleh Gambar 2.3 hanya memiliki 1 level pembelajar saja. Masing-masing algoritma tunggal memiliki akurasi prediksinya masing-masing. Setelah melewati setiap algoritma tunggal, maka diproses di bagian *voting* untuk menentukan model terbaik, akurasi paling tinggi yang akan digunakan untuk tahap selanjutnya.

2.4. Imbalanced Dataset

Dataset tidak seimbang merupakan keadaan dimana suatu dataset memiliki distribusi kelas yang tidak seimbang. Pada *dataset* tidak seimbang jumlah kelas data (*instance*) yang satu lebih sedikit atau lebih banyak dibanding dengan jumlah kelas lainnya (Ali *et al.*, 2019). Tabel 2.4 menunjukkan kategori untuk *imbalanced dataset*.

Tabel 2.5 Kategori Imbalanced Dataset

(Sumber: <https://developers.google.com/machine-learning/data-prep/construct/sampling-splitting/imbalanced-data>)

Derajat ketidakseimbangan	Proporsi dari kelas minoritas
Lemah	20-40% dari keseluruhan dataset
Sedang	1-20% dari keseluruhan dataset
ekstrim	<1% dari keseluruhan dataset

2.4.1. Synthetic Minority Over-sampling Technique

Synthetic Minority Over-sampling Technique (SMOTE) adalah metode *over-sampling* di mana data di kelas minoritas diproduksi menggunakan data sintetis yang berasal dari replikasi data di kelas minoritas (Sutoyo and Fadlurrahman, 2020). *Over-sampling* di SMOTE mengambil *instance* dari kelas minoritas dan kemudian mencari k-NN dari setiap *instance*, kemudian menghasilkan *instance* sintetis alih-alih mereplikasi *instance* kelas minoritas (Jishan *et al.*, 2015). Oleh karena itu, dapat menghindari masalah *overfitting* yang berlebihan.

2.5. Confusion Matrix

Evaluasi model digunakan untuk melihat seberapa baik metode mengklasifikasikan data. Tabel 2.5 menunjukkan *confusion matrix* yang merupakan metode paling umum untuk digunakan mengevaluasi sebuah model. Kinerja model dilihat dengan menggunakan data uji. Matriks konfusi memiliki dua dimensi yang membandingkan kelas sebenarnya dan kelas yang diprediksi.

Tabel 2.6 Confusion Matrix

		Nilai sebenarnya	
		True Positive	True Negative
Nilai Prediksi	Pred. Positive	TP	FP
	Pred. Negative	FN	TN

Dimana TP merupakan *True Positive*, FP merupakan *False Positive*, FN merupakan *False Negative* dan TN merupakan *True Negative*. Berikut merupakan rumus-rumus yang biasa digunakan pada tahap evaluasi:

$$\text{Akurasi} = \frac{\text{Total Prediksi yang benar}}{\text{Total Prediksi}} \dots\dots\dots (2.6)$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP+FN}} \dots\dots\dots (2.7)$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP+FP}} \dots\dots\dots (2.8)$$

$$\text{F1 - Score} = \frac{2 \times \text{recall} \times \text{precision}}{\text{precision+recall}} \dots\dots\dots (2.9)$$

2.6. Penelitian Terdahulu

Tabel 2.6 merupakan tabel yang memaparkan penelitian terdahulu yang dipakai sebagai acuan untuk menunjang penelitian ini.

Tabel 2.6 Penelitian Terdahulu

No.	Konten	Deskripsi
1	Judul Paper	Sentiment Analysis of Ojek Online User Satisfaction Based on the Naïve Bayes and Net Brand Reputation Method
	Penulis	Alam Rahmatulloh, Rahmi Nur Shofa, Irfan Darmawan, dan Ardiansah
	Tahun	2021
	Objek	Ojek Online
	Metode	Naïve Bayes dan NBR
	Evaluasi	Akurasi
	Hasil	Naïve Bayes menghasilkan nilai akurasi sebesar 99,80% untuk Gojek dan 99,90% untuk Grab. Sedangkan hasil perhitungan NBR menunjukkan bahwa perusahaan Grab memiliki skor kepuasan pengguna yang lebih baik daripada Gojek.
2	Judul Paper	Twitter Sentiment Analysis towards COVID-19 Vaccines in the Philippines Using Naïve Bayes
	Penulis	Charlyn Villavicencio, Julio Jerison Macrohon. Alphonse Inbaraj, Jyh-Horng Jeng and Jer-Guang Hsieh
	Tahun	2021
	Objek	Vaksin Covid-19
	Metode	Naïve Bayes
	Evaluasi	Akurasi

	Hasil	Penelitian ini memperlihatkan bahwa mayoritas atau 83% dari twit di Filipina adalah positif dan antusias dengan ide vaksinasi, sedangkan 9% netral dan 8% sentimen negatif. Data diolah terlebih dahulu menggunakan beberapa teknik NLP, dan model pengklasifikasi berhasil dikembangkan menggunakan algoritma klasifikasi Naïve Bayes dengan akurasi 81,77% melalui operator RapidMiner
3	Judul Paper	Naive Bayes Classifier on Twitter Sentiment Analysis BPJS of HEALTH
	Penulis	Sepyan Purnama Kristanto, Junaedi Adi Prasetyo, Edwin Pramana
	Tahun	2019
	Objek	BPJS Kesehatan
	Metode	Naïve Bayes
	Evaluasi	Akurasi
	Hasil	Hasil yang diperoleh adalah mendapatkan akurasi sebesar 90%. Berdasarkan hasil pengujian akurasi klasifikasi yang dilakukan dengan metode Naïve Bayes serta analisis respon komunitas twit terkait beberapa topik kritis BPJS kesehatan memperoleh respon 70% bernilai negatif.
4	Judul Paper	An Evaluation of SVM and Naive Bayes with SMOTE on Sentiment Analysis Data Set
	Penulis	Andrew Christian Flores, Rogelyn I. Icoy, Christine F. Peña, Ken D. Gorro
	Tahun	2018
	Objek	K-12 Program di Filipina
	Metode	SVM, NBM, SMOTE
	Evaluasi	Akurasi

	Hasil	Akurasi SVM cenderung lebih tinggi daripada akurasi NBM. Penggunaan SMOTE ditambah 10-fold cross validation memiliki akurasi yang lebih tinggi (84.68%) daripada penggunaan SMOTE ditambah pembagian data 70:30 (82.22%).
5	Judul Paper	Sentiment Analysis On E-Sports for Education Curriculum Using Naive Bayes and Support Vector Machine
	Penulis	Rian Ardianto, Tri Rivanie, Yuris Alkhalifi, Fitra Septia Nugraha, Windu Gata
	Tahun	2020
	Objek	E-sport untuk kurikulum pendidikan
	Metode	Naïve Bayes dan SVM
	Evaluasi	Akurasi dan AUC
	Hasil	Perbandingan dua algoritma menghasilkan prediksi yang diperoleh bahwa algoritma Naïve Bayes dengan SMOTE mendapatkan nilai akurasi 70,32%, dan nilai AUC 0,954. Sedangkan Support Vector Machine dengan SMOTE mendapatkan nilai akurasi 66,92% dan nilai AUC 0,832.
6	Judul Paper	Extraction Opinion of social media in Higher Education Using Sentiment Analysis
	Penulis	Thomas E Tarigan, Robby C Buwono, Sri Redjeki
	Tahun	2019
	Objek	Pendidikan Tinggi
	Metode	Naïve Bayes
	Evaluasi	Akurasi
	Hasil	Hasil dari penelitian ini menghasilkan sebuah sistem yang dapat mengklasifikasikan sentimen secara otomatis dengan hasil pengujian sentimen sebesar 75%, beberapa

		kendala dalam pengolahan twit secara real-time seperti twit duplikat (spam twits), struktur bahasa Indonesia yang cukup kompleks dan beragam.
7	Judul Paper	Sentiment Analysis on Covid19 Vaccines in Indonesia: From The Perspective of Sinovac and Pfizer
	Penulis	Deden Ade Nurdeni, Indra Budi, Aris Budi Santoso
	Tahun	2021
	Objek	Vaksin Covid-19: Sinovac dan Pfize
	Metode	NB, SVM, RF
	Evaluasi	akurasi
	Hasil	Hasilnya menunjukkan bahwa 77% Twit menunjukkan segmen positif, sementara 19% mewakili negatif, dan 4% tampaknya netral untuk Sinovac. Dari sudut pandang Pfizer, hasilnya masing-masing adalah 81%, 17%, dan 3% untuk positif, negatif, dan netral. Dalam hal evaluasi kinerja model, dengan validasi silang 10 kali lipat, akurasi rata-rata tertinggi dalam dataset Sinovac adalah Support Vector Machine dengan akurasi 85%. Selanjutnya, classifier Support Vector Machine memiliki nilai akurasi yang unggul 78% dalam dataset Pfizer dibandingkan classifier lainnya.
8	Judul Paper	Sentiment Analysis about E-Commerce from Twits Using Decision Tree, K-Nearest Neighbor, and Naïve Bayes
	Penulis	Achmad Bayhaqy, Sfenrianto Sfenrianto, Kaman Nainggolan, Emil R. Kaburuan
	Tahun	2018
	Objek	E-commerce
	Metode	Decision Tree, K-Nearest Neighbor, dan Naïve Bayes
	Evaluasi	Akurasi, Presisi, Recall

	Hasil	Hasil tertinggi dari penelitian ini adalah pendekatan Naïve Bayes dengan akurasi 77%, presisi 88,50% dan recall 64%.
9	Judul Paper	Sentiment Analysis Using Naive Bayes Algorithm Of The Data Crawler: Twitter
	Penulis	Meylan Wongkar, Apriandy Angdresey
	Tahun	2019
	Objek	Pemilihan Presiden
	Metode	NB, SVM, KNN
	Evaluasi	Akurasi
	Hasil	Dari hasil percobaan dapat diketahui bahwa metode Naïve Bayes memiliki tingkat akurasi yang lebih baik (yaitu 80,90%) dibandingkan dengan menggunakan metode lain, seperti KNN yang hanya memiliki tingkat akurasi 75,58% dan tingkat akurasi menggunakan SVM yaitu 63,99%.
10	Judul Paper	Implementation of Stacking Ensemble Classifier for Multi-class Classification of COVID-19 Vaccines Topics on Twitter
	Penulis	Rama Jayapermana, Aradea, Neng Ika Kurniati
	Tahun	2022
	Objek	Vaksin Covid-19
	Metode	Stacking Ensemble Classifier
	Evaluasi	Akurasi, Presisi, Recall, f1-score
	Hasil	Berdasarkan evaluasi, model Stacking Ensemble Classifier yang diusulkan menunjukkan akurasi 86%, presisi 85%, recall 86%, dan f1-score 85%.

2.7. Penelitian Terdekat

Tabel 2.7 memaparkan Penelitian yang terdekat dengan Penelitian ini.

Tabel 2.8 Penelitian Terdekat

	Penelitian Sebelumnya	Penelitian yang Diusulkan
Judul Paper/ Penelitian	Implementation of Stacking Ensemble Classifier for Multi-class Classification of COVID-19 Vaccines Topics on Twitter	Implementasi <i>Ensemble Machine Learning Classifier</i> Dan <i>Synthetic Minority Over-Sampling Technique</i> Untuk Analisis Sentimen Terhadap <i>Sustainable Development Goals</i> Di Indonesia
Penulis	Rama Jayapermana, Aradea, Neng Ika Kurniati	Nur Fajar, Acep Irham Gufroni, Irani Hoeronis
Tahun	2021	2022
Objek	Vaksin Covid-19	SDGs
Sentimen	Positif, Negatif, Netral	Positif, Negatif, Netral
Metode	Stacking Ensemble Classifier	Ensemble Machine Learning Classifier dan SMOTE
Evaluasi	Akurasi, Presisi, Recall, f1-score	Akurasi, Presisi, Recall, f1-score

2.8. Matriks Ruang Lingkup Penelitian

Table 2.9 memaparkan ruang lingkup pada penelitian ini dan perbandingannya dengan penelitian terkait sebelumnya.

Tabel 2.8 Matriks Ruang Lingkup

			Penelitian Sebelumnya	Penelitian yang Diusulkan
Judul Paper/ Penelitian			Implementation of Stacking Ensemble Classifier for Multi-class Classification of COVID-19 Vaccines Topics on Twitter	Implementasi <i>Ensemble Machine Learning Classifier</i> Dan <i>Synthetic Minority Over-Sampling Technique</i> Untuk Analisis Sentimen Terhadap <i>Sustainable Development Goals</i> Di Indonesia
Lingkup Penelitian	Metode	NB	X	✓
		SVM	✓	✓
		KNN	X	✓
		RF	✓	X
		LR	✓	X
		EMLC	✓	✓
	Sentimen	Positif	✓	✓
		Neutral	✓	✓
		Negatif	✓	✓
	Objek	Vaksin Covid-19	✓	✓
		SDGs	X	✓
	Hasil/ Evaluasi	Akurasi	✓	✓
		Presisi	✓	✓
Recall		✓	✓	
F1-score		✓	✓	
Penulis			Rama Jayapermana, Aradea, Neng Ika Kurniati	Nur Fajar, Acep Irham Gufroni, Irani Hoeronis
Tahun			2021	2022

2.9. Keterbaruan Penelitian

Berdasarkan tabel 2.9, perbedaan penelitian ini dengan penelitian sebelumnya terletak pada pengimplementasian SMOTE dan kombinasi algoritma tunggal pada *Ensemble Machine Learning Classifier* untuk meningkatkan akurasi dalam analisis sentimen. SDGs belum ada yang melakukan analisis sentimen secara keseluruhan, penelitian pada umumnya mengambil satu topik yang memiliki hubungan dengan satu tujuan pada SDGs, semisal pendidikan, kesehatan, ekonomi dan layanan publik saja. Maka dari itu, dalam penelitian ini peneliti akan melakukan analisis sentimen terhadap SDGs secara keseluruhan berdasarkan twit pengguna internet di Indonesia dengan mengimplementasikan SMOTE dan *Ensemble Machine Learning Classifier*.