

ANALISIS SENTIMEN TERHADAP KOMENTAR PADA MEDIA SOSIAL TIKTOK YANG BERPOTENSI MENYEBABKAN DEPRESI MENGGUNAKAN METODE *NAIVE BAYES*

Dinda Franciska Mey Dina¹⁾, Tining Haryanti²⁾, Muhamad Amirul Haq³⁾

^{1,2,3)} Program Studi Informatika, Fakultas Teknik, Universitas Muhammadiyah Surabaya
Jl Sutorejo No. 59, Surabaya

Email : dinda.franciska.mey-2021@ft.um-surabaya.ac.id¹⁾

Abstrak

Media sosial saat ini memiliki peran penting dalam perkembangan sebuah bisnis dan sebuah organisasi. Dari laporan America Psychiatric Association (APA) bahwa Media sosial seperti TikTok dapat mempengaruhi persepsi orang terhadap orang lain. TikTok saat ini menjadi tren dikalangan anak-anak sampai orang dewasa. Organisasi Kesehatan Dunia (WHO) menghasilkan data yang menunjukkan bahwa 1 dari 7 remaja berusia antara 10 hingga 19 tahun, yang merupakan sebagian besar pengguna TikTok, sedang berjuang dengan tantangan kesehatan mental, seperti depresi, kecemasan, dan gangguan perilaku. Komentar dalam media sosial TikTok dapat berupa komentar Positif dan Negatif, dalam jurnal Paedagogy tahun 2022 menjelaskan kecenderungan orang lebih sering berkomentar negatif hingga dapat menyerang si pemilik konten. Namun, akan sangat sulit untuk mendeteksi emosi atau sentimen dari pengguna TikTok dengan jumlah yang sangat banyak. Penelitian ini bertujuan untuk melakukan analisis sentimen yaitu untuk menganalisis komentar yang dalam tiktok yang dapat menyebabkan depresi, dengan mengklasifikasikan menjadi 2 yaitu sentimen positif, dan sentimen negatif. penelitian ini menggunakan metode Naive Bayes untuk menentukan nilai presisi, akurasi, dan F1-Score. Penelitian ini juga bertujuan agar pengguna lebih bijak sana dalam bersosial media terutama media sosial TikTok.

Kata kunci: Komentar, Naive Bayes, Sentimen Analisis

Abstract

Social media currently plays an important role in the development of a business and an organisation. From the America Psychiatric Association (APA) report that social media such as TikTok can affect people's perceptions of others. TikTok is currently a trend among children to adults. The World Health Organization (WHO) produced data showing that 1 in 7 adolescents between the ages of 10 and 19, who make up the majority of TikTok users, are struggling with mental health challenges, such as depression, anxiety, and behavioural disorders. Comments on TikTok social media can be in the form of Positive and Negative comments, in the journal Paedagogy in 2022 explained the tendency of people to comment more often negatively until they can attack the content owner. However, it will be very difficult to detect the emotions or sentiments of a very large number of TikTok users. This research aims to conduct sentiment analysis, namely to analyse comments in TikTok that can cause depression, by classifying into 2 namely positive sentiment, and negative sentiment. this research uses the Naive Bayes method to determine the precision, accuracy, and F1-Score values. This research also aims to make users wiser in social media, especially TikTok social media

Keywords : Comment, Naive Bayes, Sentiment Analysis

1. Pendahuluan

Dalam sebuah penelitian terdapat lebih dari 80% peserta melaporkan dampak yang sering terhadap penggunaan media sosial. Hasilnya mengungkapkan bahwa 48,3% ditemukan mengalami depresi, 22,6% menderita kecemasan, dan 19,4% mengalami keduanya [1] Serta merujuk pada penelitian Mardiana dan Maryana (2024) yang menemukan bahwa 60-70% remaja yang menggunakan TikTok dalam durasi tinggi cenderung mengalami peningkatan tingkat stres, depresi dan gangguan kecemasan [2] Media Sosial adalah seperangkat alat komunikasi dan kolaborasi baru yang memungkinkan terjadinya berbagai jenis interaksi yang sebelumnya tidak tersedia bagi orang awam [3] yang memungkinkan pengguna untuk berinteraksi secara langsung atau tertunda, dengan audiens yang luas sehingga banyak menimbulkan persepsi.

Menurut laporan We Are Social, pada Januari 2024 TikTok di Indonesia bisa menjangkau sekitar 126,83 juta pengguna, angka tersebut menjadikan Indonesia sebagai negara dengan jangkauan media sosial TikTok terluas ke-2 di dunia [4] Sebanyak 34,9 persen Gen Z dan Milenial di Indonesia mengalami darurat

kesehatan mental, hampir 96 persen remaja dan dewasa muda di Indonesia mengalami gejala kecemasan dan 88 persen di antaranya mengalami gejala depresi, penggunaan media sosial Tik Tok pada remaja bisa berdampak terhadap kesehatan mental remaja Berdasarkan prevalensi menurut Kementerian Kesehatan (Kemenkes) tahun 2021 di Indonesia sekitar 20% penduduk Indonesia berpotensi mengalami masalah kesehatan mental [5] Analisis sentimen merupakan salah satu studi komputasi yang bertujuan untuk menganalisis pendapat serta emosi seseorang yang dapat diekspresikan ke dalam teks [6] Sikap atau emosi tersebut dapat berupa komentar serta penilaian yang dapat dijadikan bahan evaluasi. Tujuan analisis sentimen adalah untuk mengidentifikasi pola dalam teks, kalimat dalam komentar dengan mendeteksi polaritas dalam dokumen.

Pada penelitian ini analisis sentimen untuk mengidentifikasi kecenderungan opini pada suatu masalah, apakah opini tersebut cenderung positif atau negatif, analisis sentimen terhadap komentar dapat memberikan wawasan secara real-time dan relevan [7] Naive Bayes merupakan salah satu pengklasifikasi statistik, dimana pengklasifikasi ini dapat memprediksi probabilitas keanggotaan kelas suatu data tuple yang akan masuk ke dalam kelas tertentu, sesuai dengan perhitungan probabilitas [8] menghitung probabilitas dengan syarat bahwa kelas keputusan adalah benar, mengingat vektor informasi obyek. Dari latar belakang diatas, maka dalam penelitian ini, penulis membuat pembahasan lebih dalam terkait “Analisis Sentimen Terhadap Komentar pada Media Sosial TikTok yang Berpotensi Menyebabkan Depresi Menggunakan Metode Naive Bayes” dimana `algoritma ini dapat membantu membedakan manakah komentar yang termasuk dalam klasifikasi komentar negatif dan positif dari media sosial TikTok, klasifikasi tersebut dapat membantu mengidentifikasi perasaan atau opini dengan lebih akurat

2. Dasar teori

Beberapa dasar teori yang digunakan dalam penelitian ini .

2.1. TikTok

Media sosial adalah media online, dimana para penggunanya dapat dengan mudah berpartisipasi, berbagi, dan menciptakan isi konten. Media sosial memungkinkan siapa saja untuk berpartisipasi serta memberi kontribusi atau feedback secara terbuka, memberi komentar dan membagi informasi dalam waktu yang cepat dan tidak terbatas dengan bantuan internet [9] Indonesia saat ini menduduki urutan ke-2 pengguna TikTok terbanyak di seluruh dunia. TikTok adalah aplikasi media sosial yang paling banyak diunduh di awal tahun 2022 [10] dari Gen Z, milenial hingga baby boomer semua orang beralih ke TikTok untuk bersenang-senang. Pengguna TikTok dapat membuat konten berupa video yang dapat menarik pengguna lain untuk menonton dan menyukai konten tersebut.

2.2. Analisis Sentimen

Analisis sentimen merupakan salah contoh bidang dari Neural Network Processing (NLP). Natural Language Processing (NLP) merupakan salah satu bidang ilmu Kecerdasan Buatan (Artificial Intelligence) yang mempelajari tentang interaksi antara computer dan bahasa alami manusia [11] Analisis Sentimen adalah proses komputasi menggunakan teknik analisis teks maupun gambar dalam bentuk data tekstual yang diolah dengan cara memahami dan mengelompokkan emosional menjadi sentimen positif dan sentimen negatif. Analisis sentimen tidak digunakan untuk memenuhi kebutuhan pribadi namun juga kelompok maupun organisasi. Analisis sentimen dapat digunakan juga untuk social media monitoring, brand monitoring, customer feedback, customer service, dan market research [12] Dalam melakukan analisis sentimen terbagi menjadi lima proses atau langkah yaitu crawling data, pre-processing, feature selection, classification, dan evaluation.

2.3. Depresi

Depresi adalah salah satu bentuk kondisi kesehatan mental yang dialami banyak orang dan sering kali muncul berbarengan dengan kecemasan [13] Kementerian Kesehatan baru saja mengungkapkan hasil skrining gejala depresi pada mahasiswa program pendidikan dokter spesialis atau PPDS. Data berdasarkan Patient Health Questionnaire-9 (PHQ-9) menunjukkan sekitar 22,4 persen mahasiswa

terdeteksi mengalami gejala depresi dan sekitar 3,3 persen memiliki ide bunuh diri/melukai diri sendiri [14]

2.4. Naive Bayes

Naïve bayes merupakan salah satu algoritma dari machine learning. Dalam perkembangannya Naïve Bayes termasuk kedalam golongan supervised learning yaitu salah satu machine learning yang membutuhkan sampel sebagai data latih yang memiliki label. Supervised Learning dikelompokkan menjadi dua yaitu klasifikasi dan regresi. Klasifikasi merupakan salah satu proses pada yang bertujuan untuk menemukan pola yang berharga dari data yang berukuran relatif besar hingga sangat besar [15] Dalam teori statistik dan probabilitas, teorema Bayes (juga dikenal sebagai aturan Bayes) adalah rumus matematika yang digunakan untuk menentukan probabilitas bersyarat suatu peristiwa. Intinya, teorema Bayes menggambarkan probabilitas suatu peristiwa berdasarkan pengetahuan sebelumnya tentang kondisi yang mungkin relevan dengan peristiwa tersebut.

$$P(Y|X) = P(x|y)(Y) P(X) \quad (1)$$

3. Metodologi Penelitian

Dalam tahapan penelitian ini akan menguraikan langkah-langkah kritis yang akan diambil dalam rangka melakukan penelitian tentang analisis sentimen komentar pada media sosial TikTok dengan menggunakan metode *Naive Bayes*. Melalui pendekatan sistematis ini, bertujuan untuk memberikan gambaran yang jelas tentang bagaimana data dikumpulkan, diproses, serta bagaimana model analisis sentimen diterapkan dan diuji.

3.1. Studi Literature

Pada tahap ini dilakukan studi literature dengan cara mempelajari membaca, mengutip dan membuat catatan yang bersumber pada bahan Pustaka seperti buku, skripsi, jurnal, serta referensi lainnya yang mendukung penelitian terkait Analisis Sentimen menggunakan Metode Naïve Bayes.

3.2. Identifikasi dan Perumusan Masalah

Dalam penelitian ini perumusan masalah fokus kepada komentar dalam platform media sosial Tiktok, yang dimana platform ini masih jarang dijadikan objek penelitian. Sedangkan pengguna TikTok dari tahun 2021 hingga 2024 selalu mengalami peningkatan hingga saat ini mencapai 126 juta baik dari usia anak-anak hingga dewasa. Dalam penelitian yang dipublikasikan oleh jurnal Psychiatry Research pada Juli 2023 menemukan, remaja yang kecanduan TikTok mempunyai kondisi kesehatan mental yang buruk [16]

3.3. Pengumpulan Data

Data yang digunakan dalam penelitian ini adalah kumpulan komentar dari berbagai macam konten random atau acak dalam media sosial TikTok dengan rentang waktu Januari 2023 hingga September 2024 dengan menggunakan bahasa Indonesia, karena sesuai dengan pendahuluan bahwa Indonesia adalah negara ke-2 pengguna TikTok terbanyak di dunia serta tingkat depresi yang tinggi. Dataset dalam penelitian ini berjumlah 1000 data. Data didapatkan dari crawler menggunakan website exportcomments. Berikut alur scrapping data



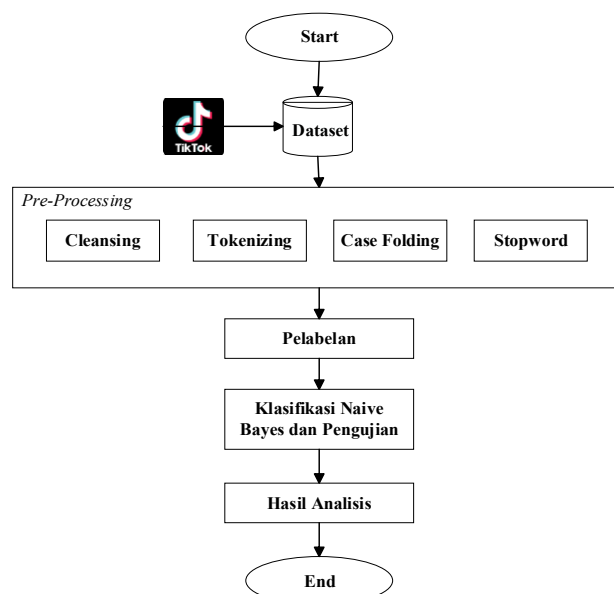
Gambar 1. Alur Scraping Data

Unique ID	Name	Date	Likes	Comment
donnadnlsongwmd	donnadnlsongwmd	03/06/24 10.52.40	133	@candy?) @Nancy
_silvilstr6	silvilstr	03/06/24 10.52.54	47	secapek ini aja di bilang Maghrib apalagi gua
bbyshampy	Anak sapi	03/06/24 10.54.33	7	yantiii cakep bgt busett
goj0osat0ru	piw	03/06/24 10.54.33	5	anjai masih sepi
cancerrrr_1171	hasbiah	03/06/24 10.54.51	7	aur auran ini mahhh
coffechill	ais	03/06/24 10.54.56	120	yg komen magrib pasti tengah malem ya?
lupanma55	Produk Madura	03/06/24 10.55.16	6	komen ke 31 kk
ahciichaadi	mochi RemajaJompo	03/06/24 10.55.20	22	ga sabar mau nonton live kamu uti
piraa.03	??	03/06/24 10.55.51	128	apasi maghrib? tan skin bagus loh mana cantik? lagi
ayumi_zha	AyumiZha	03/06/24 10.57.04	5	cakep banget tiili info bantuan umkm
_vanalex	lily of the valley	03/06/24 10.57.55	29	ntah knp ga bosan aj liat mukanya
Jjsgu	moza	03/06/24 10.58.02	85	jadi diri sendiri itu nyaman ya ti
anddiati	Ati Diati	03/06/24 10.59.08	16	mau pantun sih.. tapi udah cakep duluan..
ketanhitamz	Anelinsi	03/06/24 11.03.51	6	menyala booi Kaya raya
danialsyah_27	xyyy	03/06/24 11.07.27	20	magrib amat
berrymvffin	ButterCOokies	03/06/24 11.11.19	10	Pesona cewe tanskin menyalaaa. gue aja berusaha semaksimal mungkin njirr..
iefarafaell	barafa's	03/06/24 11.17.53	321	smoga anak ini diberi kesehatan.rizki lancar ya Allah .km dia menafkahi anak yatim piatu
xyyydaaa	@fyrr.d4	03/06/24 11.19.47	22	cantik nya habis ketemu bunda citra yaa tii
denoch_02thamrin	Denoch EgaWatie	03/06/24 11.20.52	19	gn blsi hate koment guys... ayooookk support nak cantik ini
zahrarrzkl	zazar	03/06/24 11.23.41	9	bjrrrrr CANTEKKKK bngett ckkkk

Gambar 2. Hasil Scraping Data

3.4. Pengolahan Data

Pengolahan data dimulai dari tahapan pra-proses yaitu dilakukan dengan cara membersihkan dokumen dari kata yang tidak diperlukan untuk mengurangi noise, menyeragamkan bentuk huruf serta penghapusan angka dan tanda baca, mengubah dari kalimat menjadi kata-kata, melakukan proses pemilahan kata yang merupakan kata penghubung dan mengambil kata dasar dari sebuah kata yang memiliki imbuhan.



Gambar 3. Alur Pengolahan Data

3.5. Evaluasi

Dalam evaluasi akan meliputi Akurasi, Presisi, Recall, dan *Score F-1*. Data untuk mengevaluasi didapatkan dari Confusion Matrix yaitu tabel yang digunakan untuk mengevaluasi kinerja model klasifikasi, dimana setiap baris mewakili kelas sebenarnya dan setiap kolom mewakili kelas prediksi.

Tabel 1. Confusion Matrix

	Depresi	Non-Depresi
Depresi	True Positive (TP)	False Negatif (FN)
Non-Depresi	False Positive (FP)	True Negatif (TN)

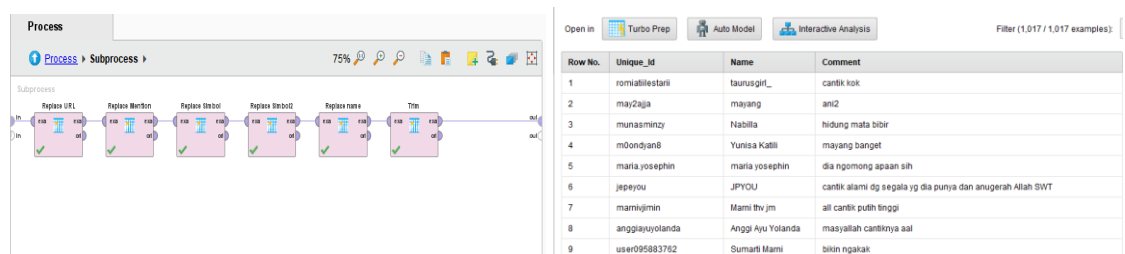
Presisi dan Recall dapat diberi nilai dalam bentuk angka dengan perhitungan presentase (1-100%) atau dengan bilangan 1-10. Sistem rekomendasi akan dianggap baik jika Presisi dan Recall-nya tinggi. Kurva ROC menunjukkan akurasi dan membandingkan klasifikasi secara visual. ROC sendiri mengekspresikan *confusion matriks*. ROC ialah grafik dua dimensi yang dimana jika *false positive* sebagai garis horizontal dan *true positive* sebagai garis vertikal.

4. Pengujian dan Pembahasan

Dalam penelitian menggunakan data-data komentar yang berasal dari TikTok dengan berbagai macam konten *Random*. Total Data yang diambil selama periode yang telah ditentukan sebanyak 1000 komentar TikTok. Dalam melakukan *crawling* data, peneliti menggunakan sebuah website *exportcoment* dengan meng-copy URL konten, yang kemudian disimpan dalam bentuk file format excel. Yang kemudian dilakukan pre processing berikut beberapa tahapannya menggunakan software Rapid Miner.

1. Cleansing

Pada *cleansing* adalah tahapan untuk membersihkan dokumen dari kata yang tidak dibutuhkan atau diperlukan guna mengurangi noise dalam data. Kata yang dihilangkan dalam *cleansing* yaitu hastag (#), URL atau link, tanda baca, symbol serta username.

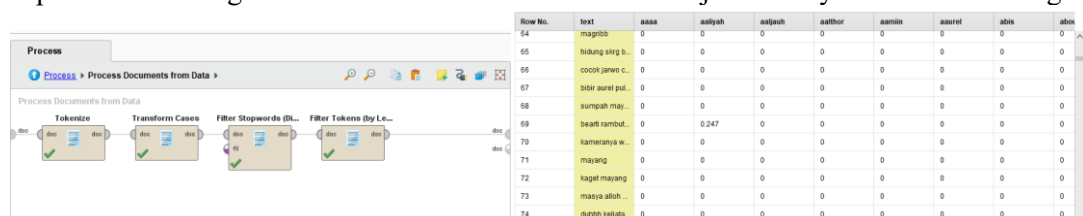


Row No.	Unique_id	Name	Comment
1	romalilestari	taurusgirl_	cantik kok
2	may2aja	mayang	am2
3	munasminz	Nabilla	hidung mata bibir
4	m0ndyan8	Yunisa Kabil	mayang banget
5	maria.yosephin	maria yosephin	dia ngomong apaan sih
6	jepayou	JPYOU	cantik alami dg segala yg dia punya dan anugerah Allah SWT
7	mamiqimin	Mami thw jm	all cantik puhi tinggi
8	anggiayulanda	Anggi Ayu Yolanda	masyallah cantiknya aal
9	user09583762	Sumarti Mami	bikin ngakak

Gambar 4. Proses cleansing dan hasil

2. Tokenizing

Setelah proses *cleansing*, tahapan selanjutnya yaitu *tokenizing* dimana pada tahap ini akan dilakukan proses pengubahan dari kalimat menjadi sebuah kata atau frasa yang mempunyai tujuan untuk mengetahui asal mula atau munculnya kata tersebut. Proses tersebut disebut juga dengan segmentasi teks. Proses ini dilakukan dengan menggunakan operator Proses Dokumen From Data dengan RapidMiner dimana akan menggunakan vektorisasi TF-IDF, proses operator ini akan berisi beberangkaian proses pre-processing dari Case Folding dan Stopword Removal. Hasil dari proses tokenizing ini berhasil memecah seluruh teks menjadi sebanyak 2.027 atribut regular.



Row No.	text	aaaa	aaliyah	aejah	aauthor	aamin	aareel	abis	abok
64	magrib	0	0	0	0	0	0	0	0
65	hidung srg b.	0	0	0	0	0	0	0	0
66	cocok jawo c.	0	0	0	0	0	0	0	0
67	bbir auel pul.	0	0	0	0	0	0	0	0
68	sumpah may.	0	0	0	0	0	0	0	0
69	beati rambut.	0	0.247	0	0	0	0	0	0
70	kameranya w.	0	0	0	0	0	0	0	0
71	mayang	0	0	0	0	0	0	0	0
72	kaget mayang	0	0	0	0	0	0	0	0
73	masya alah	0	0	0	0	0	0	0	0
74	dutuh kallata	0	0	0	0	0	0	0	0

Gambar 5. Proses dan Hasil Tokenizing

3. Case Folding

Tahap Case Folding ini merupakan penyeragaman bentuk huruf, dari huruf besar menjadi kecil atau *lower case*. Hal ini bertujuan untuk memudahkan pembacaan serta untuk pemerataan data di dalam dataset.

Row No.	text	aaa	aaayh
14	hidung	0	0
15	hidung lucu	0	0
16	alys cantik bgt	0	0
17	idung gemoy	0	0
18	mulut lambenemut	0	0
19	idung bbr	0	0
20	mulut mukanya mungil awet muda	0	0
21	adik kakak mayangmrip bngt	0	0
22	jane ngomong suka	0	0
23	mayang ngomong	0	0
24	makin temel cantikan mayang	0	0
25	aduhh cantik nyahidung mulutnya cakepawajanya imutttt pakuanya kalem banteladegga pulih aris wok anai dila	0	0
26	emang cocok bali mata	0	0
27	mayang cuantik temyata	0	0

Gambar 6. Hasil Case Folding

4. Stop Removal

Tahapan terakhir yaitu *stopword removal*, yaitu proses penghapusan kata-kata yang tidak mempengaruhi sentiment seperti kata konjungsi, preposisi, serta artikel. Dalam penelitian ini, kamus yang dipakai untuk *stopword* adalah *stopword* berbahasa Indonesia yang berasal dari Kaggle [17]. dalam proses ini dilakukan *filtering* kata berdasarkan jumlah karakter sehingga kata yang dihasilkan dataset tidak memiliki kata yang terlalu Panjang maupun pendek. Filter yang digunakan adalah penghapusan kata kecuali kata dengan minimal 4 karakter dan maksimal 25 karakter.

Row No.	text
29	lupa pale filter
30	hidungnya cantik cocok bibinya cantik
31	cantikan mayang
32	hidung hidung
33	wajah thorng jode kali
34	dihatawan orang tep cantik
35	toniq
36	kayak hamba myang anainya dirid persis banget kayak anak
37	bibir idung atah idung
38	pipi hidung bbr gemoy
39	fana alhur bal koman
40	hidungnya lucu
41	hami comen dahidung suami
42	bbr insecure trng hidung

Gambar 7. Hasil Stop Removal

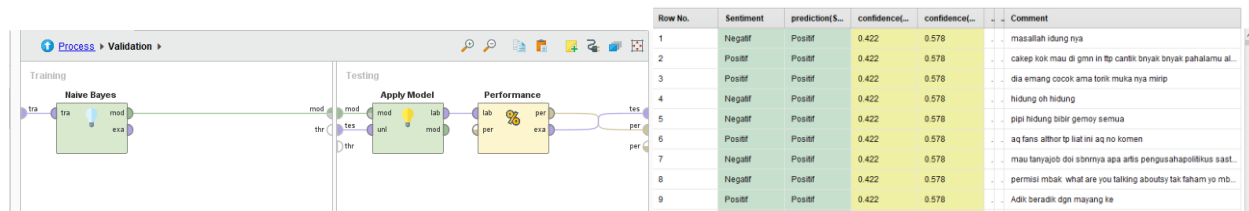
Total hasil data dalam dataset yang baru menjadi sebanyak 1.017 data, 2 atribut special, dan 2.027 atribut regular. Hal ini dikarenakan telah dilakukannya *pre-processing* yang dimana membersihkan data dari *noise* yang tidak diperlukan dalam penelitian serta data-data yang tidak relevan.

Tahapan selanjutnya setelah dilakukan proses pre-processing maka akan masuk ketahapan pelabelan. Pelabelan dilakukan untuk menentukan data tersebut masuk dalam kategori atau golongan sentiment positif, dan negatif. Pelabelan ini juga memiliki tujuan lain ialah untuk menentukan data latih yang dipakai untuk pengujian terhadap dataset. Pelabelan secara manual dilakukan dengan memberikan pelabelan kelas sentiment melalui pendapat sendiri atau ahli bahasa.

Unique_id	Name	Comment	Sentiment
1	romiatilestari	cantik kok	Positif
2	may2ajja	ani2	Negatif
3	munasminzy	hidung mata bibir	Negatif
4	m0ondyan8	mayang banget	Positif
5	maria.yosephin	dia ngomong apaan sih	Negatif
6	jepeyou	cantik alami dg segala yg dia punya dan anugerah	Positif
7	marnivjimin	all cantik putih tinggi	Positif
8	angglayuyolanda	masyallah cantiknya aal	Positif
9	user095883762	bikin ngakak	Negatif
10	sri.widyawati75	pengen ktwa	Positif
11	alhusnahanza	masalah idung nya	Negatif
12	uyun.afda	mslah apa dengan hidungnya yg penting gak opal	Positif
13	nyampahaja11	cakep kok mau di gmn in ttp cantik bnyak bnyak p	Positif

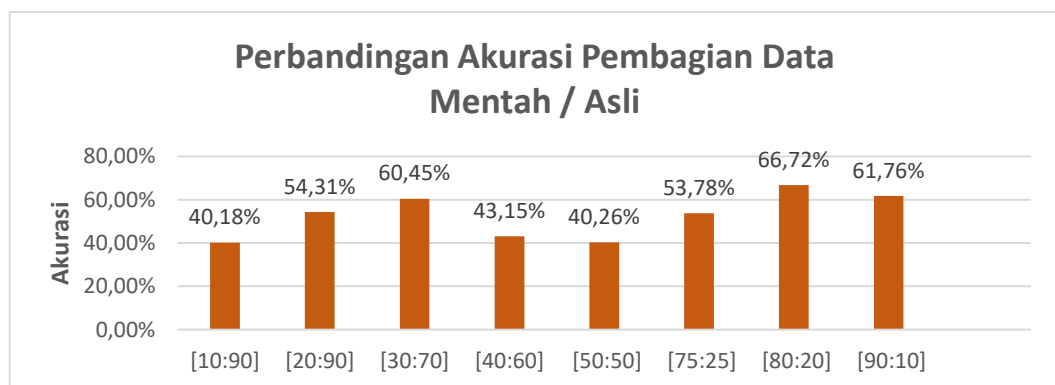
Gambar 8. Hasil Pelabelan

Tahap pengujian data ini, dataset yang telah dilakukan pre-processing serta pelabelan akan masuk dalam tahapan pengujian data. Tahapan ini diawali dengan melakukan split data atau membagi dataset yaitu data training dan data testing. Dalam pengujian data ini juga memakai k-fold cross validation untuk memperkuat hasil dari setiap pembagian data. Karena dalam penelitian ini nantinya akan membagi data set menjadi 10%, 20%, 30%, 40%, 50%, 75%, 80%, dan 90%. Data set yang telah dibagi menjadi data training dan data testing akan dievaluasi kinerja menggunakan algoritma Naïve Bayes Classifier yang akan menghasilkan estimasi akurasi, presisi, recall, dan score-F1.

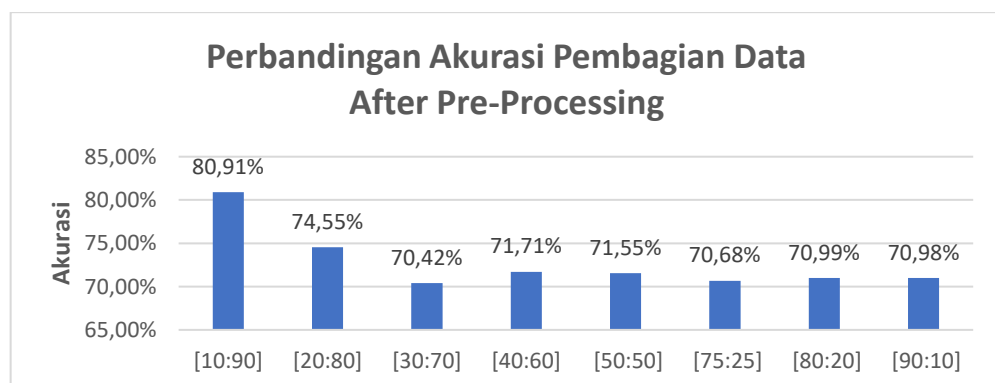


Gambar 9. Proses Klasifikasi dan hasil dengan Metode Naïve Bayes

Data akan dilakukan beberapa pengujian untuk menentukan seberapa akurasi dataset yang akan dibagi menjadi data *training* dan data *testing/uji*. evaluasi performa dilakukan dengan menggunakan *confusion matrix* untuk mengetahui nilai akurasi. Berikut tabel perbedaan hasil uji data sebelum dan sesudah dilakukan pre-processing.



Gambar 10. Perbandingan Akurasi Data Mentah / Sebelum pre-processing



Gambar 11. Hasil Perbandingan setelah Pre-processing

Dari gambar hasil evaluasi pada data setelah dilakukan pre-processing menunjukkan bahwa akurasi tertinggi terjadi pada skenario 10% atau (10:90) yaitu sebesar 80,91%. Namun, hasil ini tidak sesuai dengan teori secara umum (generalization), di mana seharusnya akurasi yang tinggi terjadi ketika data pelatihan lebih banyak dan pembagian data lebih seimbang. Sehingga, terjadi overfitting pada data training dengan hanya 10%

data sebagai pelatihan atau training, model cenderung menghafal pola sederhana pada data tersebut tanpa terlalu banyak menangkap kompleksitas. Hal ini dapat menyebabkan hasil yang tampak "bagus" pada data uji tetapi tidak mencerminkan kinerja sebenarnya. Serta jika dataset memiliki ketidakseimbangan pada label (misalnya, lebih banyak data positif atau negatif), model dapat terlihat "akurat" karena bias terhadap kelas mayoritas.

Dalam pengujian, sebaiknya pembagian data training dan data testing dipilih berdasarkan keseimbangan antara kemampuan model untuk belajar pola yang cukup dan kemampuan evaluasi yang representatif. Berdasarkan praktik umum dan hasil penelitian yang sering digunakan, 50% data untuk training dan 50% data untuk testing atau uji adalah pilihan yang baik. Pembagian ini memberikan cukup data untuk melatih model sambil menyisakan cukup data untuk mengevaluasi performa model secara akurat. skenario dengan (50:50) atau 50% menghasilkan akurasi 71.55% dengan hasil presisi 77.31%, recall 42.28%, dan F1-score 68.16%.

Dari data evaluasi diatas dapat diuraikan bahwa melakukan pre-processing sangat mempengaruhi hasil nilai keakuratan atau keakurasian data. Hasil evaluasi pada data mentah mendapat nilai akurasi tertinggi pada percobaan ke-3 dengan 80% data training dan 20% testing/uji sebesar 66.72%. Sedangkan hasil evaluasi data yang telah melalui pre-processing memiliki nilai akurasi yang jauh tinggi dan akurat. Nilai akurasi tertinggi dan tidak overfitting dan sesuai dengan teori terjadi pada evaluasi percobaan data yang telah melalui pre-processing seperti cleansing, tokenizing, case folding, dan stop removal pada skenario pertama dengan data training 50% dan data uji 50% dengan hasil sebesar 71.55%. Dimana hasil akurasi tersebut memiliki selisih akurasi sebesar 4.48% nilai yang cukup tinggi dalam analisis sebuah data.

5. Kesimpulan

Sesuai rumusan masalah dan hasil penelitian yang telah dilakukan maka dapat ditarik kesimpulan sebagai berikut :

1. Hasil penelitian menunjukkan bahwa analisis sentimen menggunakan metode Naïve Bayes Classifier melibatkan tahapan pre-processing, seperti cleansing, tokenizing, case folding, dan stopword removal. Pada tahap evaluasi, terdapat perbedaan akurasi sebesar 4,48% antara data mentah dan data yang telah melalui pre-processing. Data yang telah diproses mencapai akurasi 71,55% dalam skenario pengujian dengan 50% data, dengan presisi untuk prediksi positif sebesar 67,77% dan prediksi negatif sebesar 77,31%. Nilai recall untuk data positif adalah 82,12%, sedangkan untuk data negatif adalah 60,95%. Selain itu, nilai F1-score tercatat sebesar 68,16%. Hasil klasifikasi menunjukkan 332 data dikategorikan sebagai sentimen positif (non-depresi) dan 216 data sebagai sentimen negatif (depresi).
2. Hasil percobaan pada data mentah menunjukkan fluktuasi signifikan pada nilai akurasi akibat adanya kata-kata tidak relevan dan teks yang tidak terstruktur, sehingga sulit mengenali pola. Sebaliknya, data yang telah melalui pre-processing menunjukkan akurasi lebih tinggi dengan fluktuasi yang lebih stabil. Pre-processing berhasil meningkatkan kualitas data dengan mengurangi noise pada teks.

Penelitian ini menunjukkan bahwa metode naïve bayes efektif dalam mengukur akurasi analisis sentimen dan memberikan wawasan tentang komentar negatif beserta parameter yang mempengaruhinya. Penelitian ini juga bermanfaat untuk meningkatkan kesadaran pengguna media sosial agar lebih bijak berkomentar. Selain itu, penelitian ini diharapkan dapat memperluas pemahaman tentang analisis sentimen di media sosial, khususnya TikTok, serta memberikan panduan tentang langkah pre-processing dan skenario

data yang optimal, serta dapat digunakan untuk meningkatkan deteksi sentimen negatif dan menciptakan lingkungan digital yang lebih aman.

Daftar Pustaka

- [1] U. Zubair, M. K. Khan, and M. Albashari, "Link between excessive social media use and psychiatric disorders," Apr. 01, 2023, *Wolters Kluwer Health*. doi: 10.1097/MS9.0000000000000112.
- [2] R. Alfianto, "Candu dan Ambisi Konten FYP TikTok, 70 Persen Remaja Indonesia Mengalami Stres Karena Media Sosial," *jawapos.com*. Accessed: Jan. 07, 2025. [Online]. Available: <https://www.jawapos.com/aplikasi/015376450/candu-dan-ambisi-konten-fyp-tiktok-70-persen-remaja-indonesia-mengalami-stres-karena-media-sosial#:~:text=Unggahan%20tersebut%20merujuk%20pada%20penelitian,media%20sosial%20lainnya%20secara%20sehat.>
- [3] T. Liedfray, F. J. Waani, and J. J. Lasut, "Peran Media Sosial Dalam Mempererat Interaksi Antar Keluarga Di Desa Esandom Kecamatan Tombatu Timur Kabupaten Minahasa Tenggara," *JURNAL ILMIAH SOCIETY*, vol. 2, no. 1, 2022.
- [4] C. Annur Mutia, "Iklan TikTok di Indonesia Jangkau 126 Juta Audiens, Terbanyak ke-2 Global," *databoks*. [Online]. Available: <https://databoks.katadata.co.id/datapublish/2024/03/06/iklan-tiktok-di-indonesia-jangkau-126-juta-audiens-terbanyak-ke-2-global>
- [5] N. Ulfadilah and Y. Nurpratiwi, "Penggunaan Media Sosial Tiktok Terhadap Kesehatan Mental Remaja," Universitas Medika Suherman, Cikarang-Bekasi, 2022. Accessed: May 15, 2024. [Online]. Available: <https://repository.medikasuherman.ac.id/xmlui/handle/123456789/2855?show=full>
- [6] F. E. Kavabilla, T. Widiari, and B. Warsito, "Analisis Sentimen Pada Ulasan Aplikasi Investasi Online Ajaib Pada Google Play Menggunakan Metode Support Vector Machine Dan Maximum Entropy," *Jurnal Gaussian*, vol. 11, no. 4, pp. 542–553, Feb. 2023, doi: 10.14710/j.gauss.11.4.542-553.
- [7] A. A. S. Aziz and A. Ridok, "Analisis Sentimen Data Tweets terhadap Penanganan Covid-19 di Indonesia menggunakan Metode Naïve Bayes dan Pemilihan Kata Bersentimen menggunakan Lexicon Based," *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, vol. 6, no. 4, pp. 1865–1872, 2022, [Online]. Available: <http://j-ptiik.ub.ac.id>
- [8] A. Zainal Macfud *et al.*, "Analisis Algoritma Naive Bayes Classifier (NBC) Pada Klasifikasi Tingkat Minat Barang Di Toko Violet Cell," 2023.
- [9] C. T. Carr and R. A. Hayes, "Social Media," Binus University.
- [10] N. Sutin and R. I. Rizki, "TikTok Menjadi Trend 2022 di Platform Sosial Media," *Journal of Social and Political Science*, vol. 3, no. Januari, pp. 101–114, 2023, [Online]. Available: <https://tekno.tempo.co/read/1676881/daftar-media-sosial-terpopuler-2022-facebook->
- [11] R. Sovia, A. Syaifullah, F. Yenila, and R. Permana, "Penerapan Natural Language Processing Pada Sistem Chatbot Sebagai Helpdesk Obyek Wisata Menggunakan Metode Naïve Bayes," *Jurnal Infortech*, vol. 5, no. 2, Dec. 2023, [Online]. Available: <http://ejournal.bsi.ac.id/ejurnal/index.php/infortech210>
- [12] Y. Kurniawati, "Analisis Sentimen dan Jenisnya," Binus University. [Online]. Available: <https://sis.binus.ac.id/2023/11/24/analisis-sentimen-dan-jenisnya/>
- [13] Unicef, "Apa itu depresi?," Unicef Indonesia. Accessed: May 08, 2024. [Online]. Available: <https://www.unicef.org/indonesia/id/kesehatan-mental/artikel/depresi>
- [14] G. Pandhita, "Depresi pada Mahasiswa," *Kompas.id*. Accessed: May 10, 2024. [Online]. Available: <https://www.kompas.id/baca/opini/2024/04/18/depresi-pada-mahasiswa>
- [15] N. Yustira, D. Witarasyah, and E. Sutoyo, "Implementasi Algoritma Naive bayes Classification Untuk Klasifikasi Kelulusan Mahasiswa Tepat Waktu (Studi Kasus: Program Studi Sistem Informasi Universitas Telkom)," *Universitas Telkom Bandung*, 2021.
- [16] L. P. Amaranggana and A. N. Dzulfaroh, "Studi: Remaja yang Kecanduan TikTok Memiliki Kesehatan Mental Buruk," *Kompas.com*. Accessed: May 27, 2024. [Online]. Available: <https://www.kompas.com/tren/read/2024/02/25/140000265/studi--remaja-yang-kecanduan-tiktok-memiliki-kesehatan-mental-buruk?page=all#>
- [17] "Indonesian Stoplist," *kaggle.com*. Accessed: Dec. 19, 2024. [Online]. Available: <https://www.kaggle.com/datasets/oswinrh/indonesian-stoplist>