



Analisis Sentimen Masyarakat Pada Twitter Terhadap Kasus HIV/AIDS Di Indonesia Menggunakan Algoritma Naïve Bayes

Nurcholilah¹, Suherman², Abdul Halim Anshor³

^{1,2,3}Universitas Pelita Bangsa

Received: 28 Juni 2023

Revised: 3 Juli 2023

Accepted: 10 Juli 2023

Abstract

The high number of cases of HIV/AIDS infection which continues to increase, especially among young people or adolescents, is a serious problem. HIV data on children in Indonesia from 2020 to September 2022, there are 12,553 children aged 14 years and under who are infected with HIV with boys predominate. This shows that efforts to prevent HIV transmission from mother to child need to be increased. As social media develops, people often talk about and give their opinions through various media, one of which is social media Twitter. Opinions given by the community regarding HIV/AIDS cases varied, such as HIV screening, mandatory use of condoms so as not to get sexually transmitted diseases, and HIV testing before marriage. This research is expected to be useful in helping to conduct research on public opinion that contains sentiments, positive sentiments and negative sentiments. The method used in this study is Naive Bayes, with the process of crawling data from Twitter using Rapid Miner, preprocessing data that has been obtained from Twitter using cleansing, tokenization and filtering. For the classification process using the Naive Bayes method. The data used are tweets in Indonesian with the keywords HIV and AIDS, with a total dataset of 3.051 tweets. After going through the data selection process, the final data was 1.134 tweets with a positive sentiment of 63.49% and a negative sentiment of 36.51%. The results of testing the accuracy performance of the naïve Bayes method using RapidMiner tools obtained the best average accuracy value in the first test scenario with an average accuracy value of 59.53%, a Precision value of 68.68% and a recall value obtained of 64.89%. While the results of the accuracy performance test on Python tools using the Naïve Bayes method obtained an accuracy value of 66.00%, a precision value of 67.45% and a recall value of 94.07%

Keywords: *Sentiment Analysis, HIV/AIDS, Naïve Bayes, RapidMiner, Python.*

(*) Corresponding Author: ncholilah5@gmail.com

How to Cite: Nurcholilah, N, Suherman, S, & Anshor, A. H. (2023). Analisis Sentimen Masyarakat Pada Twitter Terhadap Kasus HIV/AIDS Di Indonesia Menggunakan Algoritma Naïve Bayes. <https://doi.org/10.5281/zenodo.8185186>

PENDAHULUAN

Human immunodeficiency virus (HIV) adalah virus yang menyerang dan merusak sistem kekebalan tubuh. Sistem kekebalan tubuh yang rusak atau lemah rentan terhadap berbagai penyakit. Sekelompok gejala penyakit yang menyerang tubuh disebut AIDS (Acquired Immune Deficiency Syndrome).

Jumlah kumulatif ODHA (kasus HIV) yang dilaporkan hingga Maret 2022 sebanyak 329.581 orang, sedangkan jumlah kumulatif kasus AIDS yang dilaporkan hingga Maret 2022 sebanyak 137.397. Jumlah kumulatif infeksi HIV di Jawa Barat hingga Oktober 2022 sebanyak 57.914 dan kumulatif kasus AIDS hingga Oktober 2022 sebanyak 12.353. Kasus HIV/AIDS di Jawa Barat terus meningkat dari tahun

ketahun, kasus pertama di Jawa Barat di temukan di Kota Bandung pada tahun 1989, dua tahun setelah kasus AIDS di Indonesia ditemukan. Data HIV pada anak di Indonesia sejak 2020 hingga September 2022, terdapat 12.553 anak usia 14 tahun ke bawah yang terjangkit HIV dengan anak laki-laki lebih dominan. Dari jumlah itu, yang sudah diobati baru sekitar 7.800 orang. Anak yang terinfeksi HIV masih ditemukan setiap tahunnya, dan sebagian besar berusia antara 25-49 dan 20-24 tahun. Hal ini menunjukkan bahwa upaya pencegahan penularan HIV dari ibu ke anak perlu diperkuat [1].

Dengan berkembangnya media sosial, masyarakat sering berekspresi dan mengungkapkan pendapatnya melalui berbagai sarana, termasuk lingkungan sosial Twitter. Pandangan masyarakat terhadap kasus HIV/AIDS beragam, seperti skrining HIV, kewajiban penggunaan kondom untuk mencegah penyakit menular seksual, dan tes HIV pranikah. Dengan memeriksa tweet ini secara lebih rinci, kami mendapatkan pendapat yang dapat kami simpulkan kasus HIV/AIDS manakah yang memiliki sentimen baik di Twitter. .

Metode yang digunakan untuk melakukan analisis sentimen adalah metode Naive Bayes Classifier. Metode tersebut adalah metode klasifikasi biner yang mengeksploitasi probabilitas statistik sederhana dengan menerapkan aturan Bayes menggunakan asumsi independen yang kuat. Metode Bayesian adalah metode analisis berdasarkan data sampel dan informasi sebelumnya. Kombinasi data sampel dengan data sebelumnya disebut probabilitas posterior. Salah satu penerapan metode Naive Bayes Classifier adalah dengan menggunakan tweet untuk menguji data. Klasifikasi dibentuk dengan mencari hyperplane terbaik yang memisahkan dua kelas dengan memaksimalkan margin. Berdasarkan latar belakang tersebut, penelitian ini melihat tweet dengan menggunakan classifier Naive Bayes yang berjudul “**Analisis Sentimen Masyarakat Pada Twitter Terhadap Kasus HIV/AIDS Di Indonesia Menggunakan Algoritma Naive Bayes**”. Penelitian ini dirancang untuk membantu meranking Twitter agar hasilnya dapat digunakan untuk menentukan dengan cepat apakah tweet tergolong negatif atau positif. .

METODE

Metode

Tujuan dari metode eksperimen adalah untuk mengetahui hubungan sebab akibat melalui perbandingan komparatif dan mengetahui seberapa besar hubungan sebab akibat tersebut. Berikut ini adalah beberapa kriteria umum untuk metode pengujian :

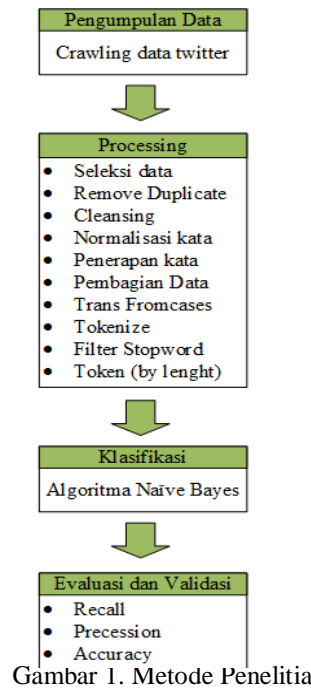
1. Pilihan masalah yang dipilih harus penting dan dapat dipecahkan .
2. Tentukan variabel dalam percobaan secara menyeluruh .
3. Jalankan pengujian yang tepat dengan rencana pengujian yang tepat .
4. Akurasi dalam observasi dan akurasi pengukuran sangat penting.
5. Jelaskan metode, bahan dan referensi yang jelas.
6. Analisis Uji Statistik.
7. Interpretasi umum.

Syarat-syarat kebaikan adalah sebagai berikut:

1. Harus bebas dari prasangka. Bertindak melawan kesalahan atau kesalahan.
2. Membuat sebuah keputusan.
3. Tentukan tujuan dengan jelas.

4. Ada banyak ruang untuk eksperimen

Metode penelitian yang digunakan penulis dalam penelitian ini digambarkan sebagai berikut :



Gambar 1. Metode Penelitian

HASIL & PEMBAHASAN

Hasil.

Hasil Pengujian RapidMiner

Hasil pengujian pertama yaitu bagaimana hasil klasifikasi sentimen yang dilakukan oleh model yang telah dibuat. Kedua bagaimana tingkat performa akurasi dari metode naïve bayes.

Hasil Pengujian Skenario Pertama

Tes dijalankan lima kali untuk rata-rata skor klasifikasi sentiment dan rata-rata kinerja akurasi metode Naive Bayes. .

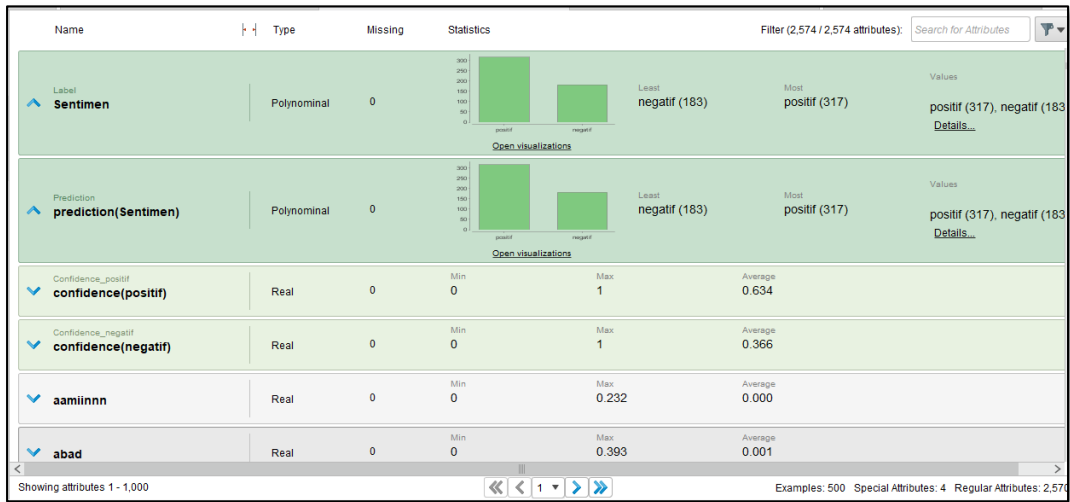
Hasil Klasifikasi Sentimen

Hasil dari proses klasifikasi sentimen pada skenario pertama yang dilakukan oleh model yang telah dibuat dapat dilihat pada tabel berikut ini :

Tabel 1. Hasil Klasifikasi Sentimen Pada Skenario Pertama

Pengujian	Hasil Klasifikasi	
	Sentimen Positif	Sentimen Negatif
1	62	38
2	63	37
3	64	36
4	58	42
5	61	39
Rata-rata	61.6	38.4

Berdasarkan tabel di atas. Terlihat rata-rata 61,6% untuk sentimen positif dan 38,4% untuk sentimen negatif. Pada skenario pengujian pertama, tweet dengan bersentimen positif lebih besar daripada tweet dengan bersentimen negatif .



Gambar 1. Contoh Statistik Hasil Klasifikasi Pada Skenario Pertama

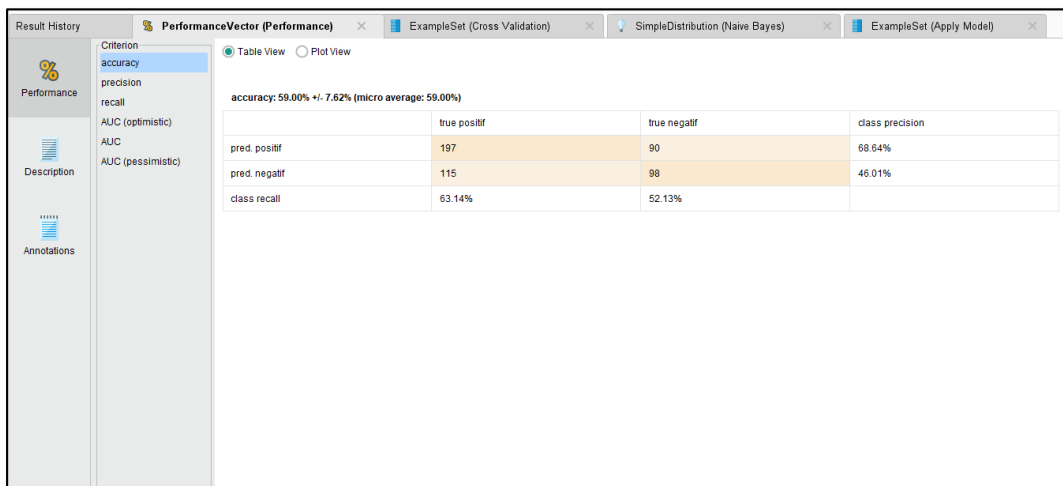
Hasil Pengujian Performa

Hasil dari pengujian performa menggunakan skenario pertama yang dilakukan sebanyak 5 kali pengujian dapat dilihat pada tabel berikut ini :

Tabel 2. Hasil Pengujian Performa Pada Skenario Pertama

Pengujian	Accuracy	Precision	Recall	AUC
1	59.00%	68.64%	63.14%	0.821
2	57.60%	68.71%	60.44%	0.813
3	62.00%	72.00%	67.08%	0.843
4	59.83%	66.23%	66.01%	0.821
5	59.20%	67.82%	67.82%	0.821
Rata-Rata	59.526%	68.68%	64.898%	0.8238

Berdasarkan tabel di atas dapat diketahui bahwa nilai akurasi tertinggi diperoleh pada pengujian 3 sebesar 62.00% dan nilai akurasi terendah diperoleh pada pengujian 2 sebesar 57.60%. Rata-rata nilai akurasi dari pengujian skenario pertama yaitu sebesar 59.526%. Nilai hasil pengujian performa tersebut diperoleh dari *Confusion Matrix* yang dapat dilihat pada contoh gambar berikut ini :



Gambar 2. Accuracy Data Testing

1. Hasil perhitungan manual Accuracy Naïve Bayes

Prediksi untuk data sentimen negatif ditemukan positif 197 dan 115 salah, sedangkan data sentimen negatif positif 98 dan salah 90. Akurasi hasil Naive Bayes dengan RapidMiner adalah 59,00%.

Hasil akurasi diperoleh dengan menjumlahkan data true positif (TP) ditambah true negatif (TN) dibagi 50% dari total data uji. Ini adalah bagaimana akurasi dihitung secara manual .

$$\begin{aligned} \text{Akurasi} &= \frac{TP + TN}{TP + TN + FP + FN} \times 100\% \\ &= \frac{197 + 98}{197 + 98 + 115 + 90} \times 100\% \\ &= \frac{295}{500} \times 100\% \\ &= \mathbf{59.00\%} \end{aligned}$$

2. Perhitungan manual *precision Naïve Bayes*

Prediksi data sentimen negatif yang diketahui positif salah 197 dan 115, sedangkan negatif positif 98 dan salah 90. Akurasi dengan Naive Bayes RapidMiner adalah 68,64%.

Hasil akurasi diperoleh dengan menjumlahkan true positive (TP) dan true negative (TN) dibagi dengan jumlah total true positive ditambah false positive (FP). Inilah cara Anda dapat melakukan perhitungan manual dengan Presisi .

$$\begin{aligned} \text{Precision} &= \frac{197}{197 + 90} \times 100\% \\ &= \frac{197}{287} \times 100\% \\ &= \mathbf{68.64\%} \end{aligned}$$

precision: 46.39% +/- 9.44% (micro average: 46.01%) (positive class: negatif)			
	true positif	true negatif	class precision
pred. positif	197	90	68.64%
pred. negatif	115	98	46.01%
class recall	63.14%	52.13%	

Gambar 3. *Precision Naïve Bayes*

3. Perhitungan Manual *Recall Naïve Bayes*

Prediksi data sentimen negatif didapat positif 197 dan 115 false, sedangkan data sentimen negatif positif 98 dan false 90. Hasil Naive Bayes dengan RapidMiner adalah 63,14%.

Hasil yang dikembalikan diperoleh dengan membagi positif sejati (TP) dengan jumlah total positif sejati (TP) dan negatif palsu (FN). Cara menghitung recall secara manual :

$$\begin{aligned} \text{Recall} &= \frac{TP}{TP + FN} \times 100\% \\ &= \frac{197}{197 + 115} \times 100\% \\ &= \frac{197}{312} \times 100\% \\ &= \mathbf{63.14\%} \end{aligned}$$

	true positif	true negatif	class precision
pred. positif	197	90	66.64%
pred. negatif	115	98	46.01%
class recall	63.14%	52.13%	

Gambar 4. Recall Naïve Bayes

Hasil Pengujian Skenario Kedua

Tes dijalankan tujuh kali untuk rata-rata skor klasifikasi sentimen dan rata-rata skor akurasi kinerja naïve Bayesian. Dalam hal ini, rata-rata digunakan sebagai pembandingan dengan rata-rata hasil skenario pengujian lainnya.

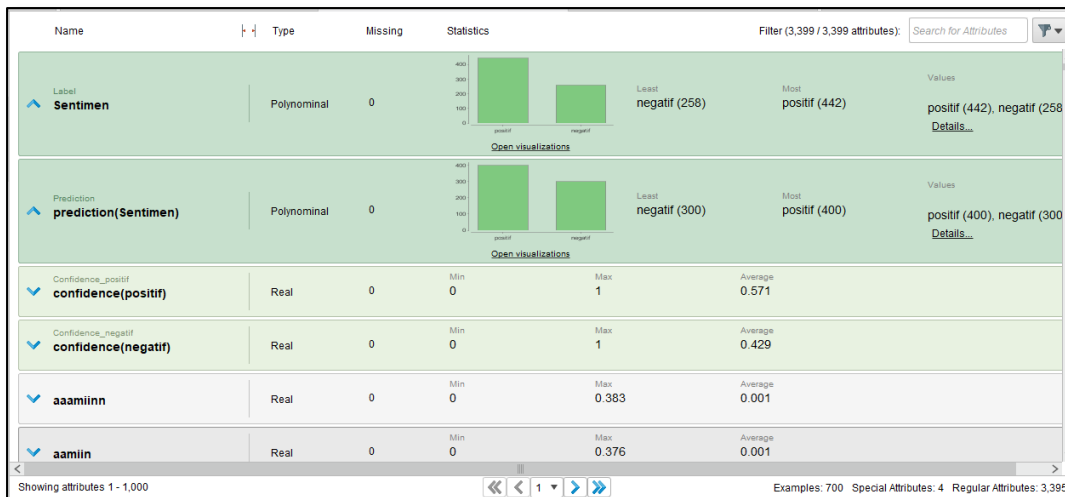
Hasil Klasifikasi Sentimen

Hasil dari proses klasifikasi sentimen pada skenario kedua yang dilakukan oleh model yang telah dibuat dapat dilihat pada tabel berikut ini :

Tabel 3. Hasil Klasifikasi Sentimen Pada Skenario Kedua

Pengujian	Hasil Klasifikasi	
	Sentimen Positif	Sentimen Negatif
1	63	37
2	61	39
3	64	36
4	65	35
5	67	33
6	64	36
7	62	38
Rata-rata	63.71	36.29

Berdasarkan tabel di atas. Terlihat nilai rata-rata sentimen positif sebesar 63,71% dan nilai rata-rata sentimen negatif sebesar 36,29%. Dengan demikian, dalam skenario pengujian, dua tweet positif lebih besar dari tweet negatif. Untuk melihat perbandingan antara hasil klasifikasi sentimen yang didapatkan secara manual dengan hasil prediksi dari model yang dibangun, dapat dilihat statistiknya seperti pada gambar berikut :



Gambar 5. Contoh Statistik Hasil Klasifikasi Pada Skenario Kedua

Hasil Pengujian Performa

Hasil dari pengujian menggunakan skenario kedua yang dilakukan sebanyak 7 kali pengujian dapat dilihat pada tabel 4.6 berikut ini :

Tabel 4. Hasil Pengujian Performa Pada Skenario Kedua

Pengujian	Accuracy	Precision	Recall	AUC
1	60.57%	70.75%	64.03%	0.837
2	55.14%	65.48%	62.25%	0.781
3	61.71%	46.76%	52.00%	0.841
4	54.57%	66.26%	60.40%	0.782
5	60.29%	69.00%	67.12%	0.822
6	57.43%	68.49%	61.74%	0.809
7	58.86%	46.38%	53.01%	0.823
Rata-Rata	58.367%	61.874%	60.078%	0.813

Berdasarkan tabel 4. dapat diketahui bahwa nilai akurasi tertinggi diperoleh pada pengujian 3 sebesar 61.71% dan nilai akurasi terendah diperoleh pada pengujian 4 sebesar 54.57%. Rata-rata nilai akurasi dari pengujian skenario kedua yaitu sebesar 58.367% Nilai hasil pengujian performa tersebut diperoleh dari *Confussion Matrix* yang dapat dilihat pada contoh gambar . berikut ini :

	true positif	true negatif	class precision
pred. positif	283	117	70.75%
pred. negatif	159	141	47.00%
class recall	64.03%	54.65%	

Gambar 6. Accuracy Data Testing

1. Hasil perhitungan manual Accuracy Naïve Bayes

Prediksi dari data sentimen negatif yang terdeteksi positif adalah 283 dan yang salah 159, sedangkan data sentiment positif yang terdeteksi negatif adalah 141 dan yang salah adalah 117. Hasil dari *Accuracy Naïve Bayes* dengan RapidMiner adalah 60.57%.

Adapun hasil *accuracy* diperoleh dengan cara menjumlah data *true positif* (TP) ditambah dengan *true negatif* (TN) dibagi dengan total jumlah data *testing* 70%. Berikut cara perhitungan manual dari *accuracy*.

$$\begin{aligned}
 \text{Akurasi} &= \frac{TP + TN}{TP + TN + FP + FN} \times 100\% \\
 &= \frac{283 + 141}{283 + 141 + 117 + 159} \times 100\% \\
 &= \frac{424}{700} \times 100\% \\
 &= \mathbf{60.57\%}
 \end{aligned}$$

2. Perhitungan manual precision Naïve Bayes

Prediksi dari data sentimen negatif yang terdeteksi positif adalah 283 dan yang salah 159, sedangkan data sentiment positif yang terdeteksi negatif adalah 141 dan yang salah adalah 117. Hasil dari *precision Naïve Bayes* dengan RapidMiner adalah 70.75%.

Adapun hasil *precision* diperoleh dengan cara menjumlah data *true positif* (TP) ditambah dengan *true negatif* (TN) dibagi dengan total jumlah data *true positif* ditambah dengan *False positif* (FP). Berikut cara perhitungan manual dari *Precision*.

$$\begin{aligned}
 \text{Precision} &= \frac{TP}{TP+FP} \times 100\% \\
 &= \frac{283}{283+117} \times 100\% \\
 &= \frac{283}{400} \times 100\% \\
 &= \mathbf{70.75\%}
 \end{aligned}$$

	true positif	true negatif	class precision
pred. positif	283	117	70.75%
pred. negatif	159	141	47.00%
class recall	64.03%	54.65%	

Gambar 7. *Precision* Naïve Bayes

3. Perhitungan Manual *Recall* Naïve Bayes

Prediksi dari data sentimen negatif yang terdeteksi positif adalah 283 dan yang salah 159, sedangkan data sentiment positif yang terdeteksi negatif adalah 141 dan yang salah adalah 117. Hasil dari *Recall Naïve Bayes* dengan RapidMiner adalah 64.03%.

Adapun hasil *Recall* diperoleh dengan cara data *true positif* (TP) dibagi dengan total jumlah data *true positif* (TP) ditambah dengan *false negatif* (FN). Berikut cara perhitungan manual dari *recall*:

$$\begin{aligned}
 \text{Recall} &= \frac{TP}{TP+FN} \times 100\% \\
 &= \frac{283}{283+159} \times 100\% \\
 &= \frac{283}{442} \times 100\% \\
 &= \mathbf{64.03\%}
 \end{aligned}$$

	true positif	true negatif	class precision
pred. positif	283	117	70.75%
pred. negatif	159	141	47.00%
class recall	64.03%	54.65%	

Gambar 8. Recall Naïve Bayes

Hasil Pengujian Skenario Ketiga

Proses pengujian dilakukan sebanyak 10 kali pengujian untuk mengetahui nilai rata-rata hasil klasifikasi sentimen dan nilai rata-rata akurasi performa naïve bayes. nilai rata-rata tersebut akan dijadikan perbandingan dengan nilai rata-rata pada hasil pengujian skenario lainnya.

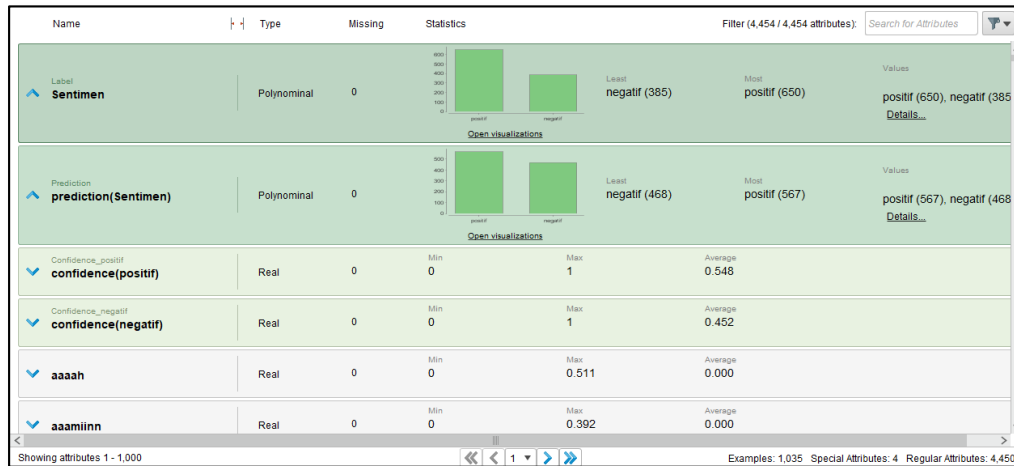
Hasil Klasifikasi Sentimen

Hasil dari proses klasifikasi sentimen pada skenario ketiga yang dilakukan oleh model yang telah dibuat dapat dilihat pada tabel 4.7 berikut ini :

Tabel 5. Hasil Klasifikasi Sentimen Pada Skenario Ketiga

Pengujian	Hasil Klasifikasi	
	Sentimen Positif	Sentimen Negatif
1	63	37
2	65	35
3	64	36
4	66	34
5	63	37
6	68	32
7	67	33
8	64	36
9	62	38
10	61	39
Rata-rata	64.3	35.7

Berdasarkan tabel di atas. dapat diketahui bahwa nilai rata-rata sentimen positif yaitu sebesar 64.3% dan nilai rata-rata sentimen negatif yaitu sebesar 35.7%. Dengan demikian pada skenario pengujian kedua *tweet* yang bersentimen positif lebih besar dibandingkan dengan *tweet* yang bersentimen negatif. Untuk melihat perbandingan hasil klasifikasi sentimen yang diperoleh secara manual dan hasil prediksi dari model yang dibuat tersebut dapat dilihat pada statistik seperti pada gambar. berikut ini :



Gambar 9. Contoh Statistik Hasil Klasifikasi Pada Skenario Ketiga

Hasil Pengujian Performa

Hasil dari pengujian menggunakan skenario ketiga yang dilakukan sebanyak 10 kali pengujian dapat dilihat pada tabel berikut ini :

Tabel 6. Hasil Pengujian Performa Pada Skenario Ketiga

Pengujian	Accuracy	Precision	Recall	AUC
1	57.00%	68.08%	59.38%	0.807
2	58.16%	44.44%	52.22%	0.817
3	60.77%	45.54%	54.22%	0.834
4	56.62%	67.79%	61.12%	0.800
5	59.13%	69.84%	61.81%	0.826
6	56.91%	42.30%	48.55%	0.804
7	57.97%	44.15%	52.36%	0.815
8	58.84%	43.49%	50.54%	0.819
9	57.10%	42.36%	51.87%	0.806
10	56.71%	68.11%	59.82%	0.803
Rata-Rata	57.921%	53.61%	55.189%	0.8131

Berdasarkan tabel di atas, dapat diketahui bahwa nilai akurasi tertinggi diperoleh pada pengujian 3 sebesar 60.77% dan nilai akurasi terendah diperoleh pada pengujian 4 sebesar 56.62%. Rata-rata nilai akurasi dari pengujian skenario kedua yaitu sebesar 57.921%. Nilai hasil pengujian performa tersebut diperoleh dari *Confussion Matrix* yang dapat dilihat pada contoh gambar 4.11. berikut ini :

	true positif	true negatif	class precision
pred. positif	386	181	68.08%
pred. negatif	264	204	43.59%
class recall	59.38%	52.99%	

Gambar 10. Accuracy Data Testing

1. Hasil perhitungan manual Accuracy Naïve Bayes

Prediksi dari data sentimen negatif yang terdeteksi positif adalah 386 dan yang salah 264, sedangkan data sentiment positif yang terdeteksi negatif adalah 204 dan yang salah adalah 181. Hasil dari Accuracy Naïve Bayes dengan RapidMiner adalah 57.00%.

Adapun hasil accuracy diperoleh dengan cara menjumlah data true positif (TP) ditambah dengan true negatif (TN) dibagi dengan total jumlah data testing 100%. Berikut cara perhitungan manual dari accuracy.

$$\begin{aligned}
 \text{Akurasi} &= \frac{TP + TN}{TP + TN + FP + FN} \times 100\% \\
 &= \frac{386 + 204}{386 + 204 + 181 + 264} \times 100\% \\
 &= \frac{590}{1035} \times 100\% \\
 &= \mathbf{57.00\%}
 \end{aligned}$$

2. Perhitungan manual precision Naïve Bayes

Prediksi dari data sentimen negatif yang terdeteksi positif adalah 386 dan yang salah 264, sedangkan data sentiment positif yang terdeteksi negatif adalah 204 dan yang salah adalah 181. Hasil dari precision Naïve Bayes dengan RapidMiner adalah 68.08%.

Adapun hasil precision diperoleh dengan cara menjumlah data true positif (TP) ditambah dengan true negatif (TN) dibagi dengan total jumlah data true positif ditambah dengan False positif (FP). Berikut cara perhitungan manual dari Precision.

$$\begin{aligned}
 \text{Precision} &= \frac{TP}{TP + FP} \times 100\% \\
 &= \frac{386}{386 + 181} \times 100\% \\
 &= \frac{386}{567} \times 100\% \\
 &= \mathbf{68.08\%}
 \end{aligned}$$

precision: 43.25% +/- 7.51% (micro average: 43.59%) (positive class: negatif)			
	true positif	true negatif	class precision
pred. positif	386	181	68.08%
pred. negatif	264	204	43.59%
class recall	59.38%	52.99%	

Gambar 11. Precision Naïve Bayes

3. Perhitungan Manual Recall Naïve Bayes

Prediksi dari data sentimen negatif yang terdeteksi positif adalah 386 dan yang salah 264, sedangkan data sentiment positif yang terdeteksi negatif adalah 204 dan yang salah adalah 181. Hasil dari *recall naïve bayes* dengan RapidMiner adalah 59.38%.

Adapun hasil *Recall* diperoleh dengan cara data *true positif* (TP) dibagi dengan total jumlah data *true positif* (TP) ditambah dengan *false negatif* (FN). Berikut cara perhitungan manual dari *recall*:

$$\begin{aligned}
 \text{Recall} &= \frac{TP}{TP+FN} \times 100\% \\
 &= \frac{386}{386+264} \times 100\% \\
 &= \frac{386}{650} \times 100\% \\
 &= 59.38\%
 \end{aligned}$$

recall: 52.94% +/- 12.75% (micro average: 52.99%) (positive class: negatif)			
	true positif	true negatif	class precision
pred. positif	386	181	68.08%
pred. negatif	264	204	43.59%
class recall	59.38%	52.99%	

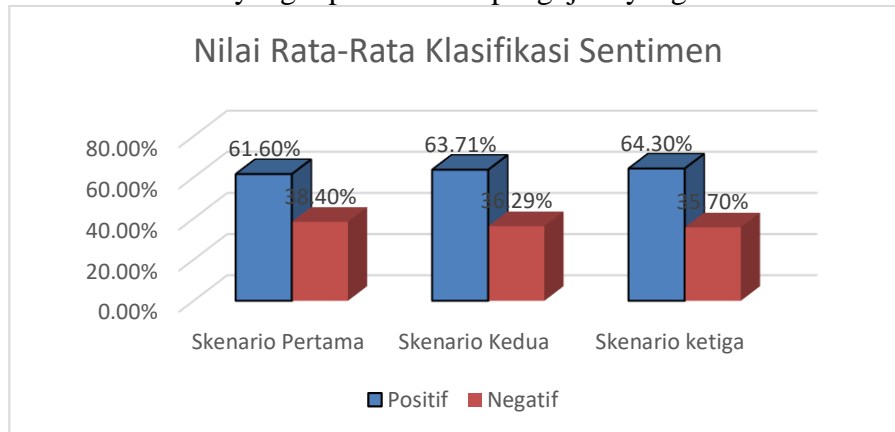
Gambar 12. Recall Naïve Bayes

Pembahasan Hasil Pengujian

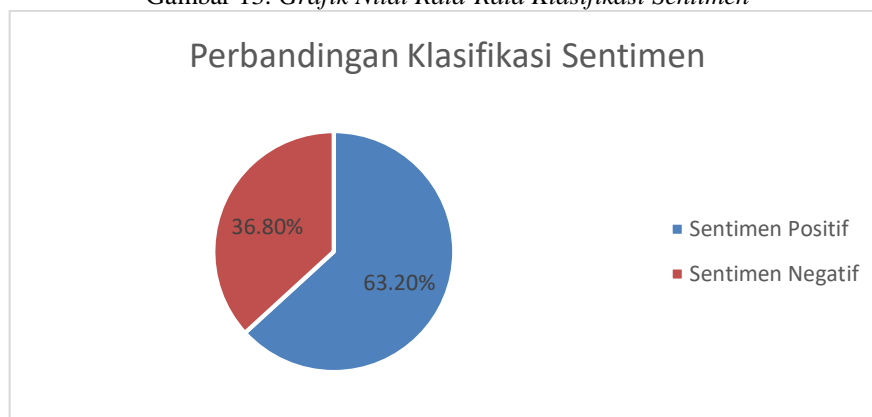
Analisis Hasil Klasifikasi Sentimen

Analisis Hasil Klasifikasi Sentimen ini bertujuan untuk mengetahui perbandingan hasil proses klasifikasi yang dilakukan oleh model yang dibuat, sehingga dapat diketahui kecenderungan sentimen masyarakat terkait penanganan

HIV/AIDS di Indonesia. Berikut adalah gambar grafik perbandingan hasil klasifikasi sentimen yang diperoleh dari pengujian yang telah dilakukan.



Gambar 13. Grafik Nilai Rata-Rata Klasifikasi Sentimen

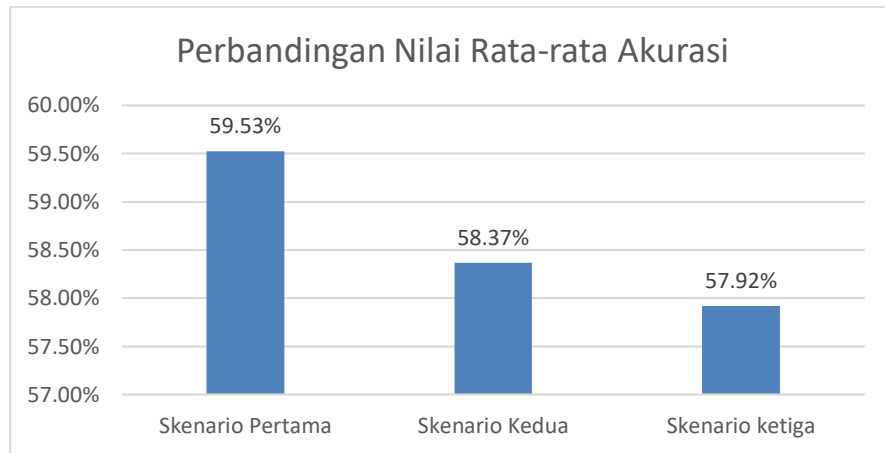


Gambar 14. Grafik Perbandingan Klasifikasi Sentimen

Berdasarkan gambar di atas, diketahui bahwa *tweet* dengan kelas sentimen positif lebih tinggi yang memiliki nilai rata-rata sebesar 63.20% dibandingkan *tweet* dengan kelas sentimen negatif yang memiliki nilai rata-rata sebesar 36.80%. Hasil klasifikasi ini selaras dengan hasil klasifikasi yang diperoleh dari penelitian Brata Mas Pintoko dan Kemas Muslim L dari Universitas Telkom Bandung pada tahun 2018 yang menghasilkan kecenderungan sentimen positif lebih besar dibandingkan dengan sentimen negatif, dengan rincian sentimen positif sebesar 88.60% dan sentimen negatif sebesar 11.40% [22].

Analisis Hasil Pengujian Performa

Analisis hasil pengujian performa ini bertujuan untuk mengetahui tingkat akurasi dari metode naïve bayes yang telah dilakukan dengan menggunakan 3 skenario pengujian dengan jumlah *dataset* yang berbeda pada tiap skenario pengujian. Hal tersebut guna membuktikan bahwa perbandingan penggunaan *dataset* yang lebih besar diharapkan dapat meningkatkan nilai akurasi dari metode naïve bayes. Berikut ini merupakan grafik perbandingan nilai rata-rata akurasi dari hasil pengujian performa naïve bayes yang dapat dilihat pada gambar 4.16.



Gambar 15. Grafik Perbandingan Nilai Rata-Rata Akurasi

Skenario pengujian ketiga dengan jumlah dataset sebanyak 1134 data *tweet*. Sedangkan nilai rata-rata akurasi terendah didapatkan pada skenario pengujian ketiga dengan jumlah dataset sebanyak 1.134 data *tweet*. Berdasarkan hasil pengujian tersebut dapat ditarik kesimpulan bahwa penggunaan dataset yang berada pada setiap skenario pengujian berpengaruh terhadap peningkatan nilai akurasi dari metode naïve bayes.

KESIMPULAN

Dari penelitian yang telah dilakukan, maka dapat ditarik kesimpulan bahwa Analisis Sentimen dapat digunakan untuk mengetahui sentimen masyarakat Indonesia terhadap *hashtag* yang sempat trending di media sosial twitter yaitu #HIV dan #AIDS. Tujuannya untuk mengetahui seberapa positif atau negatif terhadap *hashtag* tersebut. Proses pengambilan data menggunakan operator *Search* Twitter yang telah disediakan oleh perangkat lunak RapidMiner data yang berhasil di *crawling* adalah sebanyak 3.051 data. Setelah melalui proses seleksi data, didapatkan data akhir sebanyak 1.134 data *tweet* dengan jumlah sentimen positif sebanyak 63.49% dan jumlah sentimen negatif sebanyak 36,51%. Hasil pengujian performa akurasi dari metode naïve bayes menggunakan *tools* RapidMiner diperoleh nilai rata-rata akurasi terbaik pada skenario pengujian pertama dengan nilai rata-rata akurasi sebesar 59.53% , nilai *Precision* sebesar 68.68% dan nilai *recall* yang didapat sebesar 64.89%. Sedangkan hasil Pengujian Performa akurasi pada *tools* Python menggunakan metode naïve bayes diperoleh nilai akurasi sebesar 66.00% , nilai *Precision* sebesar 67.45% dan nilai *recall* yang didapat sebesar 94.07%.

REFERENCES

- [1] R. Lestari, "Tentukan Besaran Kelompok Populasi Kunci Melalui Pemetaan populasi kunci 2 – 3 tahun sekali," 2022. https://diskes.jabarprov.go.id/informasipublik/detail_berita/K2iLenFxYzRpc3R5OTEvZnJpdU1idz09 (accessed Mar. 17, 2023).
- [2] F. F. Mailo, L. Lazuardi, D. Manajemen, K. Fakultas, K. Masyarakat, and U. G. Mada, "Analisis Sentimen Data Twitter Menggunakan Metode Text Mining Tentang Masalah Obesitas di Indonesia," vol. 4, no. 1, 2019.
- [3] D. A. Ramadhan, E. Budi, and S. S. Si, "ANALISIS SENTIMEN PROGRAM ACARA DI SCTV PADA TWITTER MENGGUNAKAN

- METODE NAIVE BAYES DAN SUPPORT VECTOR MACHINE,” vol. 6, no. 2, pp. 9736–9743, 2019.
- [4] J. Jtik, J. Teknologi, S. H. Ramadhani, and M. I. Wahyudin, “Analisis Sentimen Terhadap Vaksinasi Astra Zeneca pada Twitter Menggunakan Metode Naïve Bayes dan K-NN,” vol. 6, no. 4, 2022.
- [5] M. Syarifuddin, “Inti nusa mandiri,” vol. 15, no. 1, pp. 23–28, 2020.
- [6] W. A. Prabowo and C. Wiguna, “Sistem Informasi UMKM Bengkel Berbasis Web Menggunakan Metode SCRUM,” *J. Media Inform. Budidarma*, vol. 5, no. 1, p. 149, 2021, doi: 10.30865/mib.v5i1.2604.
- [7] N. I. P. Dewi, Rafidah, and E. Yuliastuti, “Studi Literatur Faktor Yang Berhubungan Dengan Kejadian HIV/AIDS Pada Wanita Usia Subur (WUS),” *J. Inov. Penelit.*, vol. 3, no. 1, pp. 4583–4590, 2022.
- [8] F. F. Rachman and S. Pramana, “Analisis Sentimen Pro dan Kontra Masyarakat Indonesia tentang Vaksin COVID-19 pada Media Sosial Twitter,” *Heal. Inf. Manag. J.*, vol. 8, no. 2, pp. 100–109, 2020, [Online]. Available: <https://inohim.esaunggul.ac.id/index.php/INO/article/view/223/175>
- [9] E. Annisa, B. Bara, K. Amelia, R. Zahara, and G. Kartini, “Penelitian tentang Twitter,” *J. Edukasi Nonform.*, vol. 3, no. 2, pp. 1–7, 2022, [Online]. Available: <https://ummaspul.e-journal.id/JENFOL/article/view/4531/1856>
- [10] D. Darwis, E. S. Pratiwi, and A. F. O. Pasaribu, “Penerapan Algoritma Svm Untuk Analisis Sentimen Pada Data Twitter Komisi Pemberantasan Korupsi Republik Indonesia,” *Eductic - Sci. J. Informatics Educ.*, vol. 7, no. 1, pp. 1–11, 2020, doi: 10.21107/edutic.v7i1.8779.
- [11] A. Roihan, P. A. Sunarya, and A. S. Rafika, “Pemanfaatan Machine Learning dalam Berbagai Bidang: Review paper,” *IJCIT (Indonesian J. Comput. Inf. Technol.)*, vol. 5, no. 1, pp. 75–82, 2020, doi: 10.31294/ijcit.v5i1.7951.
- [12] D. Duei Putri, G. F. Nama, and W. E. Sulistiono, “Analisis Sentimen Kinerja Dewan Perwakilan Rakyat (DPR) Pada Twitter Menggunakan Metode Naive Bayes Classifier,” *J. Inform. dan Tek. Elektro Terap.*, vol. 10, no. 1, pp. 34–40, 2022, doi: 10.23960/jitet.v10i1.2262.
- [13] A. Firdaus and W. I. Firdaus, “Text Mining Dan Pola Algoritma Dalam Penyelesaian Masalah Informasi : (Sebuah Ulasan),” *J. JUPITER*, vol. 13, no. 1, p. 66, 2021.
- [14] A. Z. Malik, E. Utami, and S. Raharjo, “Analisis Sentiment Twitter Terhadap Capres Indonesia 2019 dengan Metode K-NN,” *J. Inf. Politek. Indones. Surakarta*, vol. 5, no. 2, pp. 1–7, 2019.
- [15] B. G. Sudarsono, M. I. Leo, A. Santoso, and F. Hendrawan, “Analisis Data Mining Data Netflix Menggunakan Aplikasi Rapid Miner,” *JBASE - J. Bus. Audit Inf. Syst.*, vol. 4, no. 1, pp. 13–21, 2021, doi: 10.30813/jbase.v4i1.2729.
- [16] D. Novianti, “Implementasi Algoritma Naïve Bayes Pada Data Set Hepatitis Menggunakan Rapid Miner,” *Paradig. - J. Komput. dan Inform.*, vol. 21, no. 1, pp. 49–54, 2019, doi: 10.31294/p.v21i1.4979.
- [17] Suherman and I. Muzaky, “Analisis Penjualan Barang Laris Dan Kurang Laris Terhadap Percetakan Awfa Digitl Printing Menggunakan Metode Decision Tree Dengan Optimasi Algoritma Genetika,” *J. Teknol. Pelita Bangsa*, vol. 10, no. 1, pp. 118–130, 2019.

- [18] S. Proboningrum and Acihmah Sidauruk, “Sistem Pendukung Keputusan Pemilihan Supplier Kain Dengan Metode Moora,” *JSiI (Jurnal Sist. Informasi)*, vol. 8, no. 1, pp. 43–48, 2021, doi: 10.30656/jsii.v8i1.3073.
- [19] D. Karyawan, M. Algoritma, I. Romli, and A. T. Zy, “Penentuan Jadwal Overtime Dengan Klasifikasi,” vol. 4, no. September, pp. 694–702, 2020.
- [20] R. Kausar *et al.*, “Penerapan Metode Naive Bayes Untuk Penilaian Ujian Kenaikan Tingkat Olahraga Karate,” vol. 07, pp. 178–186, 2022.
- [21] W. P. Nurmayanti, “Penerapan Naive Bayes dalam Mengklasifikasikan Masyarakat Miskin di Desa Lepak,” *Geodika J. Kaji. Ilmu dan Pendidik. Geogr.*, vol. 5, no. 1, pp. 123–132, 2021, doi: 10.29408/geodika.v5i1.3430.
- [22] B. M. Pintoko and K. M. L., “Analisis Sentimen Jasa Transportasi Online pada Twitter Menggunakan Metode Naive Bayes Classifier,” *e-Proceeding Eng.*, vol. 5, no. 3, pp. 8121–8130, 2018..