

ANALISIS SENTIMEN PUBLIK PENGGUNA TWITTER TERHADAP KANKER SERVIKS DI INDONESIA MENGGUNAKAN ALGORITMA NAIVE BAYES CLASSIFIER

Umi Faddillah¹, Ipin Sugiyarto², Lady Agustin Fitriana³

^{1,3} Sistem Informasi, Institut Universitas Bina Sarana Informatika

² Sistem Informasi, Universitas Nusa Mandiri

Jalan Kramat Raya, Jakarta Pusat, Indonesia

umi.umf@bsi.ac.id

ABSTRAK

Media sosial, khususnya Twitter, telah menjadi bagian dari kehidupan sehari-hari sebagai sumber informasi serta sebagai indikator opini publik terhadap berbagai isu yang sedang ramai diperbincangkan. Salah satu topik yang mendapat perhatian adalah kanker serviks, yang menurut data Kementerian Kesehatan RI merupakan penyakit mematikan kedua setelah kanker payudara. Perbincangan mengenai kanker serviks di Twitter mencerminkan tingkat pemahaman dan kesadaran masyarakat terhadap penyakit ini, sehingga penting untuk dianalisis lebih lanjut. Penelitian ini bertujuan untuk menganalisis sentimen publik terhadap kanker serviks di Indonesia menggunakan algoritma Naïve Bayes Classifier sebagai model klasifikasi sentimen. Data yang digunakan berupa 325 tweet terkait kanker serviks yang dikumpulkan selama lima tahun terakhir (2020–2024). Metode yang digunakan dalam penelitian ini meliputi pengumpulan data tweet, pemrosesan teks, penerapan algoritma Naïve Bayes Classifier, serta evaluasi akurasi model. Hasil penelitian menunjukkan bahwa model klasifikasi mencapai tingkat akurasi yang bervariasi setiap tahunnya, yaitu 0,76 pada tahun 2020 dengan 48 tweet positif, 0,73 pada tahun 2021 dengan 45 tweet positif, 0,76 pada tahun 2022 dengan 51 tweet positif, 0,88 pada tahun 2023 dengan 53 tweet positif, dan 0,81 pada tahun 2024 dengan 71 tweet positif. Berdasarkan hasil tersebut, dapat disimpulkan bahwa sentimen positif terhadap kanker serviks mengalami peningkatan setiap tahunnya. Hal ini mengindikasikan meningkatnya kesadaran masyarakat terhadap penyakit ini, yang dapat menjadi dasar bagi pemerintah dan lembaga kesehatan untuk terus mengedukasi masyarakat mengenai pencegahan serta deteksi dini kanker serviks guna menekan angka kasus di Indonesia.

Kata kunci : Sentimen Analisis, Twitter, Kanker Serviks, Algoritma Naïve Bayes Classifier

1. PENDAHULUAN

Kanker serviks adalah salah satu penyakit paling mematikan yang mengancam kesehatan wanita di seluruh dunia, termasuk di Indonesia. Berdasarkan laporan *Global Cancer Observatory* (GLOBOCAN), pada tahun 2020 terdapat lebih dari 36.000 kasus baru kanker serviks di Indonesia, menjadikannya sebagai kanker kedua terbanyak yang diderita oleh wanita di negara ini [1].

Penyakit ini disebabkan oleh infeksi *Human Papilloma Virus* (HPV), terutama tipe 16 dan 18, yang dapat berkembang menjadi kanker jika tidak dideteksi dan ditangani sejak dini [2].

Meski upaya pencegahan seperti vaksinasi HPV dan pemeriksaan Pap smear telah digencarkan oleh pemerintah, tantangan terbesar tetap terletak pada rendahnya kesadaran Masyarakat [3].

Beberapa di antaranya termasuk akses terbatas terhadap layanan kesehatan yang berkualitas, kurangnya kesadaran masyarakat akan pentingnya pencegahan, dan masalah sosial-budaya yang mempengaruhi perilaku kesehatan.

Era digital, media sosial telah menjadi platform utama bagi masyarakat untuk berbagi informasi, opini, dan pengalaman terkait berbagai isu, termasuk kesehatan. Twitter, sebagai salah satu media sosial paling populer di Indonesia, menjadi sumber data yang

sangat kaya untuk memahami sentimen publik terhadap suatu topik. Analisis sentimen, yaitu proses mengidentifikasi dan mengkategorikan opini atau emosi dalam teks, telah digunakan secara luas untuk menggali persepsi masyarakat dalam berbagai bidang, termasuk isu Kesehatan [4].

Dengan memanfaatkan teknologi pengolahan bahasa alami (*Natural Language Processing/NLP*), analisis sentimen dapat memberikan wawasan mendalam tentang bagaimana masyarakat memahami dan merespons isu kanker serviks [5].

Analisis sentimen merupakan metode yang digunakan untuk mengidentifikasi dan mengkategorikan opini masyarakat menjadi sentimen positif, negatif, atau netral [6].

Salah satu algoritma yang sering digunakan dalam analisis sentimen adalah *Naive Bayes Classifier* (NBC). Algoritma ini memiliki keunggulan dalam kecepatan dan efisiensi, terutama untuk data dengan volume besar seperti Twitter [7].

Naive Bayes Classifier didasarkan pada prinsip probabilitas sederhana yang mengasumsikan independensi antara fitur, meskipun pada kenyataannya fitur-fitur tersebut sering kali saling berkaitan [8].

Beberapa penelitian sebelumnya telah menerapkan NBC dalam analisis sentimen di berbagai

konteks. Pada beberapa penelitian menganalisis sentimen pada media sosial X (Twitter) terhadap tumor jinak payudara menggunakan metode *Naïve Bayes* [9].

Hasil penelitian menunjukkan bahwa NBC mampu mengklasifikasikan sentimen dengan akurasi yang memadai. Penelitian lain melakukan analisis sentimen Twitter menggunakan metode *Naive Bayes* dengan *Relevance Frequency Feature Selection*, studi kasus opini masyarakat mengenai kebijakan *New Normal* [10].

Penelitian ini menunjukkan bahwa kombinasi NBC dengan teknik seleksi fitur dapat meningkatkan akurasi klasifikasi sentimen.

Selain itu, penelitian oleh serupa juga menganalisis sentimen publik terhadap tokoh Nadiem Makarim sebagai Menteri Pendidikan dan Kebudayaan baru melalui Twitter, menggunakan algoritma *Naive Bayes* dan *Support Vector Machine* (SVM) [11].

Hasil penelitian menunjukkan bahwa metode *Naive Bayes* menghasilkan akurasi 91,48%, *precision* 89,28%, dan *recall* 91,58%, sedangkan SVM menghasilkan akurasi 90,47%, *precision* 90,23%, dan *recall* 90,78%. Temuan ini mengindikasikan bahwa *Naive Bayes* memiliki performa yang sedikit lebih baik dibandingkan SVM dalam klasifikasi sentimen pada kasus tersebut.

Meskipun telah banyak penelitian yang menerapkan NBC dalam analisis sentimen, studi yang secara spesifik menganalisis sentimen publik pengguna Twitter terhadap kanker serviks di Indonesia masih terbatas. Oleh karena itu, penelitian ini bertujuan untuk mengisi kesenjangan tersebut dengan menganalisis sentimen publik terhadap kanker serviks di Indonesia melalui platform Twitter menggunakan Algoritma *Naive Bayes Classifier*. Diharapkan hasil penelitian ini dapat memberikan wawasan yang lebih mendalam mengenai persepsi publik dan menjadi dasar bagi perumusan strategi komunikasi kesehatan yang lebih efektif.

2. TINJAUAN PUSTAKA

2.1. Sentimen Analisis

Sentiment analysis adalah proses mengidentifikasi, mengekstrak, dan menganalisis sentimen, opini, atau emosi yang diekspresikan dalam teks. Analisis ini bertujuan untuk memahami orientasi sentimen, apakah positif, negatif, atau netral, serta menggali informasi terkait emosi yang mendasari opini [12].

Analisis sentimen merupakan cabang dari pengolahan bahasa alami (NLP) yang berfokus pada pengenalan emosi dalam teks. Teknologi ini memungkinkan pengambilan keputusan berbasis opini masyarakat yang terekam dalam berbagai media, terutama media sosial [13].

Analisis sentimen sebagai proses penggalian informasi untuk memahami sikap, emosi, atau opini

seseorang terhadap suatu topik, produk, atau layanan berdasarkan teks yang dihasilkan [14].

Jadi, dapat disimpulkan Sentiment analysis adalah proses NLP untuk mengidentifikasi, mengekstrak, dan menganalisis sentimen, opini, atau emosi dalam teks. Analisis ini bertujuan memahami orientasi sentimen (positif, negatif, atau netral) serta menggali informasi untuk mendukung pengambilan keputusan berbasis opini masyarakat di berbagai media, terutama media sosial.

2.2. Twitter

Twitter sebagai alat komunikasi sosial dengan fokus pada penyebaran informasi cepat, percakapan langsung, dan pembentukan jaringan komunitas. Fitur-fiturnya, seperti hashtag dan *retweet*, mendukung eksplorasi isu-isu terkini serta keterlibatan publik [15].

Twitter adalah media sosial yang sederhana namun efektif, memungkinkan pengguna berbagi pesan singkat dalam berbagai format. Di Indonesia, platform ini populer untuk menyampaikan opini, kritik, dan saran secara *real-time*, menjadikannya alat komunikasi yang cepat dan interaktif [16].

Dapat didefinisikan, Twitter merupakan media sosial sederhana dan efektif yang mendukung penyebaran informasi cepat, komunikasi personal dan publik, serta pembentukan komunitas. Platform ini populer untuk berbagi opini dan mengikuti tren sosial secara *real-time*.

2.3. Kanker Serviks

Kanker serviks sebagai jenis kanker yang berkembang di sel-sel leher rahim (serviks), yaitu bagian bawah rahim yang menghubungkan rahim dengan vagina. Penyebab utama kanker serviks adalah infeksi *Human Papilloma Virus* (HPV), terutama tipe onkogenik seperti HPV 16 dan HPV 18, yang dapat memicu pertumbuhan sel abnormal dan akhirnya berkembang menjadi kanker [17].

Kanker serviks merupakan salah satu penyebab utama kematian akibat kanker pada wanita di seluruh dunia. Penyakit ini sering kali tidak menunjukkan gejala pada tahap awal [18].

2.4. Algoritma Naïve Bayes Classifier

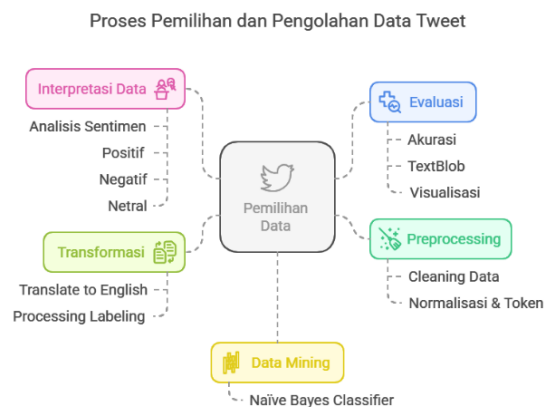
Algoritma Naive Bayes adalah metode klasifikasi yang berbasis pada teorema Bayes dengan asumsi kuat bahwa setiap fitur saling independen satu sama lain. Algoritma ini digunakan secara luas dalam tugas klasifikasi teks, seperti analisis sentimen, karena kesederhanaannya dan efisiensinya [19].

Menurut Cremer (2021), *Naive Bayes* mempengaruhi suatu fitur dalam memprediksi kelas bersifat independen terhadap fitur lainnya, meskipun dalam praktiknya asumsi ini sering kali tidak sepenuhnya valid [20].

Kumari (2021), menyebutkan bahwa meskipun asumsi "*naive*" ini tidak realistis, algoritma tetap memberikan kinerja yang baik, terutama dalam dataset yang besar [21].

3. METODE PENELITIAN

Metode penelitian ini menggunakan model *Knowledge Discovery in Database (KDD)* yang mencakup tahapan seleksi data dari cuitan Twitter, preprocessing untuk membersihkan data, transformasi ke bentuk *numerik*, *data mining* menggunakan algoritma *Naive Bayes Classifier* untuk analisis sentimen, dan evaluasi model menggunakan metrik akurasi. Adapun tahapan metodologi penelitian dijabarkan dalam Gambar 1.



Gambar 1. Metodologi Penelitian Model KDD

3.1. Pemilihan Data

Tahap ini pemilihan data dengan menggunakan data komentar bersumber dari media sosial twitter X mulai rentang 5 tahun dari 2020 sampai dengan 2024 dengan jumlah tweet 325 tweet. Di tahun 2020 terdapat 59 tweet, tahun 2021 sebanyak 64 tweet, tahun 2022 terdapat 56 tweet, tahun 2023 sebanyak 60 tweet dan tahun 2024 terdapat 86 tweet.

3.2. Preprocessing

Tahap *preprocessing* data ini pertama melakukan proses *cleaning* data dengan langkah awal *lower text* membuat seluruh text yang bercampur dengan huruf besar menjadi seragam berhuruf kecil. Selanjutnya menghapus *re-tweet* pada setiap komentar, menghapus hashtag atau tanda pagar pada setiap baris komentar, menghapus link alamat website setiap baris komentar, menghapus simbol atau karakter khusus pada setiap komentar, menghapus kata duplikat serta mengkoreksi data null.

Tahap kedua di *preprocessing* melakukan proses normalisasi data dengan menyederhanakan kata-kata menjadi kata baku, proses *stopword* menghilangkan kata penghubung, *tokenize* melakukan proses pemenggalan kata pada setiap kalimat menjadi bentuk per-satu kata dan terakhir melakukan proses stemming untuk menghapus kata imbuhan pada setiap kata dasar di masing-masing kalimat.

3.3. Transformasi

Proses transformasi ini akan merubah bentuk data menjadi bentuk atau konteks berbeda. Dataset awal yang berubah komentar dalam bahasa Indonesia

akan di ubah menjadi dalam bentuk bahasa inggris dengan tujuan memudahkan proses *labelling* data dengan bantuan NLTK (*Natural Language Toolkit*) data text berupa komentar akan lakukan *labeling* data menjadi 3 kategori label berupa komentar positif, negatif & netral.

3.4. Data Mining

Tahap data mining ini bagian penting untuk membantu menyelesaikan problem dalam bentuk algoritma yang berbasis klasifikasi. Naive bayes ini menggunakan teknik klasifikasi sesuai dengan rumus dasarnya yaitu dengan memperhitungkan probabilitas berdasarkan munculnya kata pada setiap kejadian dan terhadap seluruh kejadian. Konteks ini mampu memberikan perhitungan sesuai dengan formula matematisnya dengan konsep probabilitasnya pada setiap masing masing peluang di setiap event atau kejadiannya.

3.5. Evaluasi

Tahap ini akan membuat model visualisasi dalam bentuk *textblob* untuk memudahkan dalam menganalisa dengan menampilkan beberapa text yang sering muncul dengan karakteristik ukuran text dan warna yang menunjukkan dominasi berbeda menyesuaikan intensitas munculnya kata pada setiap event. Selain itu bentuk evaluasi data juga di visualkan dalam bentuk *bar-char* yang sangat membantu model dalam klasifikasinya sesuai dengan konteks 3 kelas komentar positif, negatif dan normal. Mempertimbangkan hasil akurasi model memastikan hasil mencapai ekspektasi positif, negatif atau netral yang sudah sesuai dengan pendekatan penilaian pada realitanya.

3.6. Interpretasi

Tahap terakhir ini pada bagian interpretasi hasil dengan memberikan pandangan atau pendapat terkait hasil olah data dari masing masing kategori komentar positif, negatif dan netral untuk dijadikan issue terkini dari masyarakat tentang mengetahui bahaya dan langkah pencegahan penyakit kanker serviks serta memberikan insight dari visualisasi data untuk membuat suatu kebijakan terhadap langkah pencegahan dan sosialisasi kepada masyarakat.

4. HASIL DAN PEMBAHASAN

Penelitian ini menggunakan *Algoritma Naive Bayes Classifier (NBC)* yang diimplementasikan dengan bahasa pemrograman Python untuk melakukan analisis sentimen. Proses dimulai dengan pemilihan data dari dataset Twitter yang relevan, kemudian dilanjutkan dengan tahapan preprocessing untuk membersihkan data dari elemen yang tidak diperlukan. Selanjutnya, dilakukan transformasi data teks menjadi format numerik yang dapat diolah oleh algoritma. Tahap data mining menggunakan NBC bertujuan untuk mengklasifikasikan data ke dalam kategori sentimen positif, negatif, atau netral. Hasilnya

dievaluasi menggunakan metrik seperti akurasi, presisi, dan recall, diikuti dengan interpretasi data untuk memberikan wawasan terkait sentimen publik.

4.1. Pemilihan Data

Sample data tweet tahun 2020		Sample data tweet tahun 2021	
0	hi butter aku mau minta alias minta up in amp.	0	skining riwayat sehat perlu deteksi sakit din
1	teman2 mohon bantu doa amponi sembuh sauda	1	imunisasi hpv imunisasi beri cegah sakit kanker
2	ibunda ressa herlambang tinggal juang lawan sa	2	ibunda ressa herlambang tinggal juang lawan sa
3	kanker payudara kanker serviks yang waspada ru	3	cegah kanker serviks kak ns
4	aku pake vaksin gardasil kita lebih baik bandi	4	akar baja herbal kalmantan sempat viral obat
5	kok takut hanti padahal tau ga ngejeks aku leb	5	kams 17 november 2022 pukul 09 00 vbo s d sel
6	yth ts ngg k oke kren pasien dgn diagnosa s	6	mis nanya siri keluarga idap kanker serviks dila
7	ihut insiden kanker serviks orang vaksin hpv	7	dokter anak atas tina tahun vaksin hpv cegah k
8	bis ngk jokes reply engak bales padahal mau m	8	sobat sehat apa kanker leher rahim kanker serviks
9	1 12 dokter rati sang kayak becak seadehat	9	hpv cegah kanker serviks ns
10	kanker serviks kanker dua perempuan takut kan	10	mjb di keluarga besar riwayat kanker bukan ser
Sample data tweet tahun 2022		Sample data tweet tahun 2023	
0	skining riwayat sehat perlu deteksi sakit din	0	rek kalo klan baca curhatan panjang kak sint
1	imunisasi hpv imunisasi beri cegah sakit kanker	1	vates tempat aia kakurahan ngeistharjo bhabin
2	ibunda ressa herlambang tinggal juang lawan sa	2	aku cari skining kanker serviks bisa home car
3	cegah kanker serviks kak ns	3	kavan kabinet kabar gembira nih mual tahun va
4	akar baja herbal kalmantan sempat viral obat	4	cmasa hari ada suhu kanker serviks vaksin hpv
5	kams 17 november 2022 pukul 09 00 vbo s d sel	5	kanker paru paru rupa jenis kanker jadi sambian
6	mis nanya siri keluarga idap kanker serviks dila	6	hi ester informasi media sosial tidak sebut im
7	dokter anak atas tina tahun vaksin hpv cegah k	7	tk kunk ugum sama tk ui ri kanker dharmas brit
8	sobat sehat apa kanker leher rahim kanker serviks	8	sender buh responden perempuan teiti kati k
9	hpv cegah kanker serviks ns	9	kanker serviks silent killer paling banyak inc
10	mjb di keluarga besar riwayat kanker bukan ser	10	kalkan covek jangan rokok beneran dulu pemah
Sample data tweet tahun 2024			
0	perempuan wajib tau kanker serviks tempat urut	0	
1	kenal tular virus hpv human papillomavirus	1	
2	1 warga negara n meninggal setiap 1 jam kanker serviks	2	
3	bukan gak sebenarnya gak kalo pas tangen kelu	3	
4	promo vaksin hpv rs adi husada kapasari sedang	4	
...		...	
81	kata dokter 5 gejala kanker serviks susah tan	81	
82	mbak uthi takut kena kanker serviks jd uthi jara	82	
83	kanker serviks rupa salah satu sakit prur vira	83	
84	curhatannya instagram final miss universe 2020	84	
85	yang tau guideline pedoman kanker serviks apa	85	

Gambar 2. Sample Data Tweet Tahun 2022-2024

Pada sample dataset tweet tersebut terdapat total 325 tweet dari mulai data tahun 2020 sampai dengan 2024. Masing-masing di tahun 2020 dataset tweet sebanyak 59 tweet kemudian setelah dilakukan *preprocessing* data menjadi 58 tweet. Di tahun 2021 data tweet sebanyak 64 tweet, setelah di *preprocessing* menjadi tetap 64 tweet. Tahun 2022 data tweet sebanyak 56 setelah melewati *preprocessing* tetap 56 tweet. Data tahun 2023 sebanyak 60 tweet setelah melewati *preprocessing* tetap 60 tweet. Data tahun 2024 sebanyak 86 tweet setelah melewati *preprocessing* tetap 86 tweet. Jadi secara keseluruhan data clean yang dijadikan sebagai data training total menjadi 324 data tweet selama jangka waktu 5 tahun seperti yang di deskripsikan pada Gambar 2.

4.2. Preprocessing

[] # Proses lower text df['clean'] = df.apply(lambda row: row['full_text'].lower(),axis=1)	
[] df[['full_text','clean']].head()	
full_text	Clean
0 PEREMPUN WAJIB TAU !! Kanker serviks menempat...	perempuan wajib tau !! kanker serviks menempat...
1 Mari Kenali Penularan Virus HPV HPV (Human p...	mari kenal penularan virus hpv hpv (human p...
2 1 warga negara n meninggal setiap 1 jam karen...	1 warga negara n meninggal setiap 1 jam karen...
3 Bukan gak boleh sebenarnya yg gak boleh itu ka...	bukan gak boleh sebenarnya yg gak boleh itu ka...
4 Promo Vaksin HPV RS Adi Husada Kapasari sedang...	promo vaksin hpv rs adi husada kapasari sedang...

Gambar 3. Proses Lower Text

Proses gambar 3, ini dilakukan untuk merubah huruf besar menjadi huruf kecil dari setiap tweet dengan mendeteksi masing masing susunan kata dalam sebuah kalimat. Data *full_text* ini merupakan data asli sebelum dilakukan proses sedangkan Clean merupakan data hasil dari *lower text* untuk membuat seluruh huruf dalam setiap tweet menjadi standar huruf kecil.

[7] #Proses Menghapus @retweet import re [8] df['clean'] = df.apply(lambda row: re.sub("@[A-Za-z0-9_]+","", row['clean']),axis=1) [9] df[['full_text','clean']]	
full_text	Clean
0 PEREMPUN WAJIB TAU !! Kanker serviks menempat...	perempuan wajib tau !! kanker serviks menempat...
1 Mari Kenali Penularan Virus HPV HPV (Human p...	mari kenal penularan virus hpv hpv (human p...
2 1 warga negara n meninggal setiap 1 jam karen...	1 warga negara n meninggal setiap 1 jam karen...
3 Bukan gak boleh sebenarnya yg gak boleh itu ka...	bukan gak boleh sebenarnya yg gak boleh itu ka...
4 Promo Vaksin HPV RS Adi Husada Kapasari sedang...	promo vaksin hpv rs adi husada kapasari sedang...
...	...
89 Kata Dokter Ini 5 Gejala Kanker Serviks yang S...	kata dokter ini 5 gejala kanker serviks yang s...
90 @ryshw_ Mbaknya uthi takut kena kanker serviks ...	mbaknya uthi takut kena kanker serviks jd uthi ...
91 Kanker serviks merupakan salah satu penyakit y...	kanker serviks merupakan salah satu penyakit y...
92 Dalam curhatannya di Instagram finalis Miss Un...	dalam curhatannya di instagram finalis miss un...
93 [cmj] ada yang tau guideline atau pedoman untuk...	[cmj] ada yang tau guideline atau pedoman untuk...

Gambar 4. Proses Menghapus Re-Tweet

Tahap gambar 4, merupakan proses menghilangkan *re-tweet* pada setiap komentar. Bagian *full_text* adalah text asli *tweet* yang belum dilakukan proses penghilangan *re-tweet* sedangkan *Clean* bagian *text* yang sudah dilakukan proses menghilangkan *re-tweet*.

# Proses menghapus hashtag df['clean'] = df.apply(lambda row: re.sub("#[A-Za-z0-9_]+","", row['clean']),axis=1) df[['full_text','clean']]	
full_text	Clean
0 PEREMPUN WAJIB TAU !! Kanker serviks menempat...	perempuan wajib tau !! kanker serviks menempat...
1 Mari Kenali Penularan Virus HPV HPV (Human p...	mari kenal penularan virus hpv hpv (human papillomavirus) merupakan virus yang dapat menyebabkan berbagai penyakit termasuk kanker serviks. vaksin hpv adalah cara terbaik untuk melindungi diri dari virus ini. segera lakukan vaksinasi dan dapatkan promo vaksin hpv. https://t.co/SF8nydAFZ
2 1 warga negara n meninggal setiap 1 jam karen...	1 warga negara n meninggal setiap 1 jam karena kanker serviks then whyyyy gak ada subsidi yg significant buat vaksin hpv!!!!.....
3 Bukan gak boleh sebenarnya yg gak boleh itu ka...	bukan gak boleh sebenarnya yg gak boleh itu kalo pas menangani keluarga berakhir subjektif. Contoh: diagnosis yg betul kanker serviks stadium 2 tp karena dmya liat kok keluarganya sakit bgt nih kayanya stadium 3 deh bukan 2 itu yg sebaiknya dihindari oleh dokter
4 Promo Vaksin HPV RS Adi Husada Kapasari sedang...	promo vaksin hpv rs adi husada kapasari sedang menyediakan promo vaksin hpv untuk pencegahan primer kanker serviks! promo berlaku dari 20 april 4 mei 2024 kuota terbatas ! *syarat dan ketentuan berlaku https://t.co/vujpzeTVzv

Gambar 5. Proses Menghilangkan Hashtag Tweet

Pada proses gambar 5, ini menghilangkan tanda hashtag (#) pada setiap komentar. Setiap kata kata dalam kalimat akan di scan dan jika terdapat tanda pagar (#) atau fungsi hashtag maka akan terhapus. Terlihat pada *full_text* adalah sumber text asli sebelum proses sedangkan Clean adalah hasil yang sudah dilakukan proses penghilangan hashtag.

# Proses Menghapus Links df['clean'] = df.apply(lambda row: re.sub("http[s]*","", row['clean']),axis=1) df[['full_text','clean']]	
full_text	Clean
0 PEREMPUN WAJIB TAU !! Kanker serviks menempat...	perempuan wajib tau !! kanker serviks menempat...
1 Mari Kenali Penularan Virus HPV HPV (Human p...	mari kenal penularan virus hpv hpv (human papillomavirus) merupakan virus yang dapat menyebabkan berbagai penyakit termasuk kanker serviks. vaksin hpv adalah cara terbaik untuk melindungi diri dari virus ini. segera lakukan vaksinasi dan dapatkan promo vaksin hpv. https://t.co/SF8nydAFZ
2 1 warga negara n meninggal setiap 1 jam karen...	1 warga negara n meninggal setiap 1 jam karena kanker serviks then whyyyy gak ada subsidi yg significant buat vaksin hpv!!!!.....
3 Bukan gak boleh sebenarnya yg gak boleh itu ka...	bukan gak boleh sebenarnya yg gak boleh itu kalo pas menangani keluarga berakhir subjektif. Contoh: diagnosis yg betul kanker serviks stadium 2 tp karena dmya liat kok keluarganya sakit bgt nih kayanya stadium 3 deh bukan 2 itu yg sebaiknya dihindari oleh dokter
4 Promo Vaksin HPV RS Adi Husada Kapasari sedang...	promo vaksin hpv rs adi husada kapasari sedang menyediakan promo vaksin hpv untuk pencegahan primer kanker serviks! promo berlaku dari 20 april 4 mei 2024 kuota terbatas ! *syarat dan ketentuan berlaku https://t.co/vujpzeTVzv
5 Keputihan Menahan Sakit Perut Walitra Sehat Wanita 100% Asli Tanpa Bahan Kimia hanya Rp112.000 - Rp194.000 di Shopee sekarang juga! https://t.co/xZBakubu	Keputihan Menahan Sakit Perut Walitra Sehat Wanita 100% Asli Tanpa Bahan Kimia hanya Rp112.000 - Rp194.000 di Shopee sekarang juga!

Gambar 6. Proses Menghilangkan Links

Gambar 6, ini akan dilakukan penghapusan sumber links pada setiap akhir kalimat yang biasanya sering disertakan. Jadi di tahap ini hanya akan meninggalkan text saja tanpa sumber links https dan sumber url lainnya. Full_text merupakan sumber text asli sedangkan Clean adalah hasil dari proses yang sudah menghilangkan sumber links.

```
#Proses hapus Filtering non-alphanumeric
df['Clean'] = df.apply(lambda row: re.sub("[^a-z0-9]", " ", row['Clean']), axis=1)
```

```
df[['full_text', 'Clean']]
df
```

index	full_text
0	PEREMPUAN WAJIB TAU !! Kanker serviks menempati urutan kedua sebagai jenis kanker yang paling banyak dialami oleh wanita di Indonesia dan 70% penderita kanker serviks terlambat dideteksi ITULAH MENGAPA PENTINGNYA VAKSIN HPV
1	Mari Kenali Penularan Virus HPV HPV (Human papillomavirus) merupakan virus yang dapat menyebabkan berbagai penyakit termasuk kanker serviks. Vaksin HPV adalah cara terbaik untuk melindungi diri dari virus ini. Segera lakukan vaksinasi dan dapatkan promo vaksin HPV. https://t.co/SF8nydAFZ
2	1 warga negara ri meninggal setiap 1 jam karena kanker serviks THEN WHY??? gak ada subsidi yg significant buat vaksin hpv!!!
3	Bukan gak boleh sebenarnya yg gak boleh itu kalo pas menangani keluarga berakhir subjektif. Contoh: diagnosis yg betul kanker serviks stadium 2 tp karena drnya liat kok keluarganya sakit bgt nih kayanya stadium 3 deh bukan 2 itu yg sebaiknya dihindari oleh dokter
4	Promo Vaksin HPV RS Adi Husada Kapasari sedang menyediakan promo vaksin HPV untuk pencegahan primer kanker serviks! Promo berlaku dari 20 April 4 Mei 2024 KUOTA TERBATAS ! sjarat dan ketentuan berlaku #promovaksin #promovaksinhpv #vaksinhpv https://t.co/Q8gy19PxeK

Gambar 8. Proses Penghilangan Karakter Non-Alphanumeric

Proses gambar 8, ini merupakan filtering karakter non-alphanumeric pada setiap kata kata di dalam text. Jika ada maka akan di hapus seperti pada bagian data Clean sudah dilakukan proses penghilangan dapat di bandingkan dengan full_text pada data sebelumnya.

```
[ ] df.shape
(94, 2)
```

```
[ ] df = df.drop_duplicates(subset=['Clean'])
```

```
[ ] df.duplicated().sum()
0
```

```
[ ] df = df.dropna()
```

```
[ ] df.isnull().sum()
full_text    0
Clean        0
dtype: int64
```

```
[ ] df.shape
(86, 2)
```

	full_text	Clean
0	PEREMPUAN WAJIB TAU !! Kanker serviks menempati...	perempuan wajib tau kanker serviks menempati...
1	Mari Kenali Penularan Virus HPV HPV (Human p...	mari kenali penularan virus hpv hpv human p...
2	1 warga negara ri meninggal setiap 1 jam karen...	1 warga negara ri meninggal setiap 1 jam karen...
3	Bukan gak boleh sebenarnya yg gak boleh itu ka...	bukan gak boleh sebenarnya yg gak boleh itu ka...
4	Promo Vaksin HPV RS Adi Husada Kapasari sedang...	promo vaksin hpv rs adi husada kapasari menyed...

Gambar 9. Hasil Proses Penghilangan Data Duplikat

Tahap gambar 9, ini merupakan proses melakukan penghilangan data yang memiliki duplikat pada komentar atau tweet. Jika terdapat duplikat maka akan di hapus tweet tersebut seperti terjadi pada proses ini di mana data awal tweet berjumlah 94 tweet menjadi 86 tweet.

```
#Normalisasi
norm = { "yg": " yang ", "tp": "Tetapi", "dm": "pesan pribadi",
        "tuk": "untuk", "alm": "Almarhum", "tb": "tinggi badan",
        "mintol": "minta tolong", "yki": "yayasan kanker indonesia",
        "ml": "making love", "utk": "untuk",
        "rscm": "rumah sakit cipto mangun kusumo",
        "udah": "sudah", "ngumpuln": "mengumpulkan",
        "ununtuk": "untuk", "ununuununuununuununtuk": "untuk" }

def normalisasi (str_text):
    for i in norm:
        str_text = str_text.replace(i, norm[i])
    return str_text

df['Clean'] = df['Clean'].apply(lambda x: normalisasi(x))
df
```

	Clean
0	perempuan wajib tahu kanker serviks menempa...
1	mari kenali penularan virus hpv hpv human p...
2	1 warga negara ri meninggal setiap 1 jam karen...
3	bukan tidak boleh sebenarnya yang tidak bole...
4	promo vaksin hpv rs adi husada kapasari sedang...
...	...
89	kata dokter ini 5 gejala kanker serviks yang s...
90	mbaknya udh takut kena kanker serviks jd udh ...
91	kanker serviks merupakan salah satu penyakit y...
92	dalam curhatannya di instagram finalis miss un...
93	ada yang tahu guideline atahu pedoman untuk ...

Gambar 10. Proses Normalisasi

Pada gambar 10, normalisasi ini beberapa kata yang tidak menggunakan kata baku seperti, “gak” menjadi “tidak”, kata “tau” menjadi “tahu” dan lainnya. Proses ini untuk memudahkan dalam menejemen kata kata yang memiliki arti sampai ke akarnya menurut bahasa Indonesia yang baik dan benar untuk keperluan library berbahasa indonesia yang nantinya akan di install pada library sastrawi untuk memudahkan proses.

```
#Stopword
import Sastrawi

from Sastrawi.StopwordRemover.StopwordRemoverFactory
import StopwordRemoverFactory, StopwordRemover, ArrayDictionary
more_stop_words = ["tidak", "dan", "yang", "atau", "dengan", "di", "lalu", "itu",
                  "kemudian", "lalu", "untuk", "karena", "sejak", "sampai",
                  "sebelum", "ketika", "setelah", "jadi", "selanjutnya",
                  "serta", "lagipula", "setelah", "sejak", "selanjutnya",
                  "tetapi", "melainkan", "sedangkan", "maupun", "ataupun",
                  "agar", "supaya", "sebab", "sehingga", "akibatnya",
                  "jika", "kalau", "jikalau", "walaupun", "meskipun", "biarpun",
                  "seperti", "sebagai", "bagaikan", "biar", "biarpun",
                  "bukannya", "melainkan", "bila"]
```

	Clean
0	perempuan wajib tahu kanker serviks menempa...
1	kenali penularan virus hpv hpv human papill...
2	1 warga negara ri meninggal 1 jam kanker servi...
3	bukan sebenarnya pas menangani keluarga bera...
4	promo vaksin hpv rs adi husada kapasari menyed...

Gambar 11. Proses stopword

Tahapan gambar 11, stopword ini setiap kata sambung seperti, “dan”, “agar”, “meskipun” dan lainnya akan di hapus sehingga hanya akan tersisah kata dasar saja yang memiliki makna. Hal ini akan memudahkan proses di mana pendeteksian kata dasar secara tepat yang memiliki arti dan maksud.

```
#Tokenize
tokenized = df['Clean'].apply(lambda x:x.split())
tokenized

0      [perempuan, wajib, tahu, kanker, serviks, mene...
1      [kenali, penularan, virus, hpv, hpv, human, pa...
2      [1, warga, negara, ri, tinggal, 1, jam, kank...
3      [bukan, sebenarnya, pas, tangan, keluarga, ...
4      [promo, vaksin, hpv, rs, adi, husada, kapasari...
...
89     [kata, dokter, 5, gejala, kanker, serviks, sss...
90     [mbaknya, udh, takut, kena, kanker, serviks, j...
91     [kanker, serviks, merupakan, salah, satu, peny...
92     [curhatannya, instagram, finalis, miss, univer...
93     [tahu, guideline, atahu, pedoman, kanker, serv...
Name: Clean, Length: 86, dtype: object
```

Gambar 12. Proses Tokenize

Pada tahapan gambar 12 ini setiap kalimat akan di pisah menurut masing masing kata pembetuknya. Teknik ini dilakukan untuk memudahkan scoring pada setiap frekuensi munculnya kata kata tersebut secara keseluruhan dokumen data text.

```
# Stemming
from Sastrawi.Stemmer.StemmerFactory import StemmerFactory

def stemming (text_cleaning):
    factory = StemmerFactory()
    stemmer = factory.create_stemmer()
    do = []
    for w in text_cleaning:
        dt = stemmer.stem(w)
        do.append(dt)
    d_clean = []
    d_clean = " ".join(do)
    print(d_clean)
    return d_clean

tokenized = tokenized.apply(stemming)

kek ga cocok menscup keluar mulu pdhl ukur xs pai
16 renal cortex 17 duodenum usus 12 jari jejenum
fun bike charity rupa wujud komitmen jalan tanggu
apresiasi lalu giat harap tingkat sadar masyarakat
moga giat tingkat sadar masyarakat bahaya kanker
tuju selenggara giat rangka ingat bulan peduli ka
virus hpv sebab kanker serviks sebut sebab kanker
fun bike charity diselenggarakan sama yayasan kank
fun bike charity rupa acara gelar posind sama yay
ini jawab anak bebas tinggi badan berat polio kan
fun bike charity cukup galang dana bantu obat kan
ikut ciri darah vagina akibat kanker serviks
fun bike charity cukup galang dana obat kanker se
ajak masyarakat peduli kanker serviks posind gela
bagi ingat bulan peduli kanker serviks fun bike c
proses jelas dokter kak buat cegah kanker serviks
acara harap tingkat sadar masyarakat kanker servi
funbikevharithy dpt tingkat sadar masyarakat kanke
fun bike charity adakan cukup galang dabs bantu o
```

Gambar 13. Proses Stemming

Tahap stemming yang dijabarkan pada gambar 13, merupakan proses mereduksi kata kata menjadi kata dasarnya misal, “dituju” menjadi “tuju”, “serupa” menjadi “rupa” dan lainnya. Proses ini membantu menemukan kata dasar sesuai dengan kamus besar bahasa Indonesia untuk memudahkan mengetahui makna dasar dari masing masing kata.

	Clean	tweet_english
0	perempuan wajib tahu kanker serviks tempat uru...	women must know cervical cancer where the two ...
1	kenal tular virus hpv hpv human papillomavirus...	know how to contract the virus hpv hpv human p...
2	1 warga negara ri tinggal 1 jam kanker serviks...	1 citizen ri stay 1 hour cervical cancer then ...
3	bukan sebenarnya pas tangan keluarga akhir sub...	it's not really a subjective final family hand...
4	promo vaksin hpv rs adi husada kapasari sedia ...	hpv vaccine promo rs adi husada cottonari avai...
...	...	...

Gambar 14. Proses Penerjemahan Bahasa Indonesia ke Bahasa Inggris

Proses gambar 14, merupakan penerjemahan bahasa Indonesia ke Inggris tujuan agar memudahkan dalam proses pembuatan label setiap komentar. Dengan menggunakan fungsi library pada translate akan memudahkan untuk menerjemahkan seluruh text ke dalam bahasa Inggris.

4.3. Transformasi

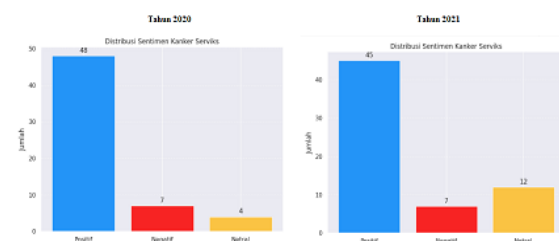
	Clean	tweet_english	Klasifikasi
0	perempuan wajib tahu kanker serviks tempat uru...	women must know cervical cancer where the two ...	Positif
1	kenal tular virus hpv hpv human papillomavirus...	know how to contract the virus hpv hpv human p...	Positif
2	1 warga negara ri tinggal 1 jam kanker serviks...	1 citizen ri stay 1 hour cervical cancer then ...	Positif
3	bukan sebenarnya pas tangan keluarga akhir sub...	it's not really a subjective final family hand...	Negatif
4	promo vaksin hpv rs adi husada kapasari sedia ...	hpv vaccine promo rs adi husada cottonari avai...	Positif
...	...	...	...
81	kata dokter 5 gejala kanker serviks sssudah t...	said the doctor 5 symptoms of cervical cancer ...	Netral
82	mbak udh takut kena kanker serviks jd udh jara...	mbak udh is afraid of getting cervical cancer ...	Negatif
83	kanker serviks rupa salah satu sakit perlu was...	cervical cancer looks like one of the sick nee...	Positif
84	curhatannya instagram final miss universe 2020...	curhatannya instagram final miss universe 2020...	Negatif
85	tahu guideline atahu pedoman kanker serviks ap...	know the guidelines or know the guidelines for...	Netral

Gambar 15. Hasil Labeling Data

Pada tahapan gambar 15, merupakan proses transformasi data text yang sebelumnya sudah di pisahkan masing masing kata dan sudah di translate tetapi belum memiliki makna berarti di bagian ini dilakukan labeling data berdasarkan masing masing konteks kata yang sudah di sesuaikan masing masing arti atau maknanya untuk mengarahkan ke klasifikasi netagif, positif atau netral. Proses ini di bantu dengan library *Natural Language Toolkit* (NLTK) untuk membantu mengarahkan ke sentiment analysis setelah seluruh proses mulai dari tokenisasi sampai dengan parsing kata selesai.

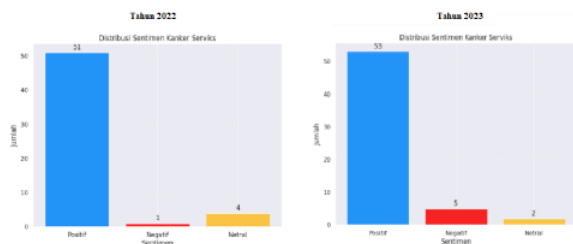
4.4. Data Mining

Tahap ini merupakan proses penambangan data dengan menggunakan model teknik algoritma *naives bayes classifier* dengan kemampuannya dalam menentukan peluang masing masing klasifikasi antara positif, negatif dan netral terhadap keseluruhan data untuk menentukan sesuai dengan kelasnya masing masing. Hasil yang di dapatkan dari proses ini bisa kita lihat visualisasi data sentimentnya pada bar-chart gambar 16.



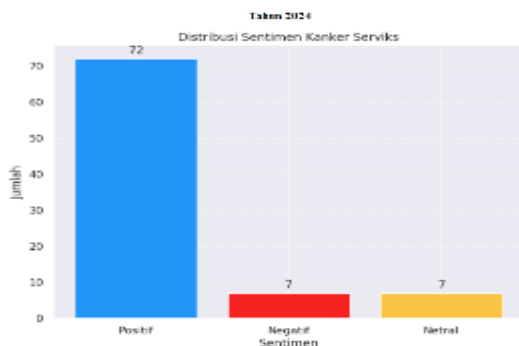
Gambar 16. Visualisasi Naives Bayes Classifier Tahun 2020 - 2021

Hasil ini didapatkan bahwa pada tahun 2020 terdapat sentimen positif sebanyak 48 tweet, sentiment negatif sebanyak 7 tweet dan sentiment netral 4 tweet. Berdasarkan hasil visual selama 2 tahun terakhir terjadi penurunan data sentiment positif dari 47 tweet menjadi 45 tweet dan terjadi kenaikan sentimen netral menjadi 12 tweet di tahun 2021.



Gambar 17. Visualisasi Naive Bayes Classifier Tahun 2022-2023

Hasil ini didapatkan bahwa pada tahun 2022 terdapat sentimen positif sebanyak 51 tweet, sentiment negatif sebanyak 1 tweet dan sentiment netral 4 tweet. Berdasarkan hasil gambar 17, selama 2 tahun terakhir terjadi kenaikan data sentiment positif dari 51 tweet menjadi 53 tweet dan terjadi kenaikan sentiment negatif menjadi 5 tweet di tahun 2023 dan penurunan jumlah tweet netral.



Gambar 18. Visualisasi Naive Bayes Classifier Tahun 2024

Hasil ini didapatkan bahwa pada tahun 2024 terdapat sentimen positif sebanyak 72 tweet, sentiment negatif sebanyak 7 tweet dan sentiment netral 7 tweet. Berdasarkan hasil gambar 18, selama 5 tahun terakhir terjadi kenaikan data sentiment positif dari 53 tweet menjadi 72 tweet dan terjadi kenaikan sentiment negatif menjadi 7 tweet di tahun 2024 dan kenaikan jumlah tweet netral sebanyak 7 tweet.

4.5. Evaluasi

Tahap ini merupakan hasil evaluasi dari algoritma naive bayes classifier terhadap model yang diterapkan pada studi kasus ini untuk mengklasifikasikan menjadi 3 luaran sentiment positif, negatif dan netral. Hasil evaluasi ini dalam ukuran evaluasi nilai akurasi dari masing masing data tweet setiap tahunnya selama jangka waktu 5 tahun mulai 2020-2024. Hasil evaluasi akurasi test didapatkan dari data training yang di ambil sampelnya secara random berdasarkan sumber dataset setiap tahunnya. Sehingga bisa dikomparasikan sebagai berikut di table 1.

Tabel 1. Summary akurasi naive bayes classifier

	Data Tweet per-tahun				
Tahun	2020	2021	2022	2023	2024
Sentimen Positif	48	45	51	53	74
Sentimen Negatif	7	7	1	5	7
Netral	4	12	4	2	5
Akurasi Test	0,762 7118	0,734 375	0,767 8571	0,8	0,802 3255
Total Data Tweet	59	64	56	60	86
Rata-rata Akurasi	0,77345388				

4.6. Interpretasi

Tahap ini untuk interpretasi data digunakan model visual word cloud sebagai indikasi konteks isu apa yang sering muncul pada setiap tahun mengenai kasus kanker serviks sebagai berikut;



Gambar 19. Visualisasi Word Cloud Data Tahun 2020-2021

Berdasarkan hasil visual gambar 19, kumpulan kata yang paling dominan terdapat kata “kanker”, “serviks”, “hpv”, “cegah”, “sebab”, “vaksin”, “perempuan” dan “ibu”. Asumsi bahwa di tahun 2020 membuktikan masyarakat mulai mencari tahu tentang penyebab kanker serviks terutama pada perempuan tentang ketakutan akan penularan virus *human papillomavirus* (hpv) dan ada pandangan mengenai pencegahan sakit akibat virus tersebut. Lalu pada tahun 2021 mulai muncul pembicaraan mengenai vaksin yang diperuntukan seorang wanita untuk para calon ibu sebagai langkah dini cegah terkena penularan hpv atau kanker serviks.



Gambar 20. Visualisasi Word Cloud Data Tahun 2022-2023

Hasil visualisasi gambar 20, pada tahun 2022 secara sebagian hampir mirip banyak isu yang senada. Uniknya di tahun ini dokter mulai memiliki pandangan vaksin terhadap anak-anak sebagai langkah dini pencegahan. Lalu di tahun 2023 juga isu mengenai vaksin masih bertahan dan pembicaraan mengenai cegah dengan mengecek rahim perempuan dan isu imunisasi pun tidak luput dari pembicaraan sebagai langkah cegah penularan ke anak-anak.



Gambar 21. Visualisasi word cloud data tahun 2024

Pada gambar 21, tahun 2024 ini sudah mulai muncul isu lain yang mendukung program yang dapat membantu pencegahan kanker serviks, mulai dari obat sampai beberapa kegiatan kegiatan positif seperti bersepeda kemudian charity dengan tujuan Indonesia peduli terhadap virus HPV penyebab dari kanker serviks serta beberapa isu juga ada dari yayasan kanker yang ikut bergerak menyuarakan untuk bantu pencegahan.

Selain pandangan terhadap banyak isu yang muncul pada setiap tahunnya dapat dihubungkan dengan peningkatan jumlah sentiment positif pada setiap tahunnya menunjukkan akan kesadaran masyarakat terhadap isu kanker serviks ini yang di buat oleh model *naive bayes classifier* dalam menentukan kelas positif, negatif dan netral dengan rata-rata persentase akurasi sekitar 0.77 % dengan jumlah data selama kurun waktu 5 tahun dari 2020 sampai 2024 sebesar 325 tweet. Skor akurasi mulai dari tahun 2020 sebesar 0,762 % dengan jumlah data tweet positif sebesar 48 dan tweet negatif 7 serta tweet netral 4 menunjukkan peningkatan skor akurasi di setiap tahunannya dan terakhir di tahun 2024 mendapatkan skor 0,802% dengan jumlah data tweet positif sebanyak 74 dan tweet negatif tetap berapa di skor 7 serta untuk tweet netral sebanyak 7, dari sini bisa dikorelasikan bahwa semakin banyak data latih yang kita gunakan maka hasil akurasi semakin membaik. Hal ini membuktikan bahwa model ini akan lebih baik hasilkan jika banyak data latih yang digunakan dengan komposisi data yang heterogen model mampu belajar lebih tepat lagi dalam melakukan klasifikasi.

5. KESIMPULAN DAN SARAN

Penelitian ini menunjukkan bahwa analisis sentimen publik terhadap kanker serviks di Indonesia menggunakan algoritma *Naive Bayes Classifier* (NBC) mampu memberikan hasil yang akurat dan

relevan. Dengan memanfaatkan data dari 325 tweet selama periode 2020 hingga 2024, model ini berhasil mengklasifikasikan sentimen ke dalam kategori positif, negatif, dan netral dengan tingkat akurasi rata-rata sebesar 77,3%. Hasil analisis memperlihatkan peningkatan jumlah sentimen positif setiap tahunnya, yang mencerminkan peningkatan kesadaran masyarakat terhadap kanker serviks. Model NBC terbukti efektif dalam pengelolaan data dengan volume besar dan menghasilkan klasifikasi yang andal, meskipun terdapat keterbatasan dalam asumsi independensi antar fitur. Selain itu, interpretasi data melalui visualisasi seperti *word cloud* juga memberikan wawasan tentang isu utama yang menjadi perhatian publik setiap tahunnya.

Untuk meningkatkan hasil penelitian di masa mendatang, disarankan mengkombinasikan algoritma *Naive Bayes Classifier* (NBC) dengan algoritma lain seperti *Support Vector Machine* (SVM) atau teknik *Deep Learning*, guna meningkatkan akurasi dan efisiensi klasifikasi sentimen. Selain itu, penggunaan dataset yang lebih besar dan beragam akan memberikan hasil analisis yang lebih representatif terhadap opini publik. Penelitian ini juga dapat menjadi dasar bagi organisasi kesehatan dalam merancang kampanye kesadaran masyarakat yang lebih efektif dengan memanfaatkan media sosial. Kampanye tersebut dapat difokuskan pada edukasi terkait pencegahan kanker serviks, seperti pentingnya vaksin HPV, deteksi dini, dan gaya hidup sehat, untuk meningkatkan kesadaran dan keterlibatan Masyarakat

DAFTAR PUSTAKA

- [1] "International Agency for Research on Cancer, "Global Cancer Observatory: Cervical Cancer." <https://gco.iarc.fr>.
- [2] J. Ferlay *et al.*, "Cancer statistics for the year 2020: An overview," *Int. J. Cancer*, vol. 149, no. 4, pp. 778–789, 2021, doi: 10.1002/ijc.33588.
- [3] "Ministry of Health Indonesia, 'Laporan Situasi Kesehatan Perempuan.'" <https://www.kemkes.go.id/>.
- [4] F. Greco, "Sentiment analysis and opinion mining," in *Elgar Encyclopedia of Technology and Politics*, 2022, pp. 105–108.
- [5] Y. Wang, J. Guo, C. Yuan, and B. Li, "Sentiment Analysis of Twitter Data," *Applied Sciences (Switzerland)*, vol. 12, no. 22. 2022, doi: 10.3390/app122211775.
- [6] H. Juwiantho *et al.*, "Sentiment Analysis Twitter Bahasa Indonesia Berbasis Word2vec Menggunakan Deep Convolutional Neural Network," *J. Teknol. Inf. dan Ilmu Komput.*, vol. 7, no. 1, pp. 181–188, 2020, doi: 10.25126/jtiik.202071758.
- [7] R. R. Putra, M. E. Johan, and E. R. Kaburuan, "A naïve bayes sentiment analysis for fintech mobile application user review in Indonesia," *International Journal of Advanced Trends in Computer Science and Engineering*, vol. 8, no. 5.

- pp. 1856–1860, 2019, doi: 10.30534/ijatcse/2019/07852019.
- [8] M. Karmagatri, I. D. Utama, H. Yogie, J. L. Chen, and H. Hasan, “NAIVE BAYES SENTIMENT ANALYSIS ON PERCEPTIONS OF HALAL CERTIFICATION: A CASE STUDY ON MIXUE INDONESIA,” *J. Theor. Appl. Inf. Technol.*, vol. 101, no. 17, pp. 6781–6789, 2023.
- [9] S. A. Lestari, “Analisis Sentimen Pada Media Sosial X (Twitter) Terhadap Tumor Jinak Payudara Menggunakan Metode Naïve Bayes,” vol. 5, no. 3, pp. 3336–3348, 2024.
- [10] K. V. S. Toy, Y. A. Sari, and I. Cholissodin, “Analisis Sentimen Twitter menggunakan Metode Naive Bayes dengan Relevance Frequency Feature Selection (Studi Kasus: Opini Masyarakat mengenai Kebijakan New Normal),” *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 5, no. 11, pp. 5068–5074, 2021, [Online]. Available: <http://j-ptiik.ub.ac.id>.
- [11] T. T. Widowati and M. Sadikin, “Analisis Sentimen Twitter terhadap Tokoh Publik dengan Algoritma Naive Bayes dan Support Vector Machine,” *Simetris J. Tek. Mesin, Elektro dan Ilmu Komput.*, vol. 11, no. 2, pp. 626–636, 2021, doi: 10.24176/simet.v11i2.4568.
- [12] L. Yang, Y. Li, J. Wang, and R. S. Sherratt, “Sentiment Analysis for E-Commerce Product Reviews in Chinese Based on Sentiment Lexicon and Deep Learning,” *IEEE Access*, vol. 8, pp. 23522–23530, 2020, doi: 10.1109/ACCESS.2020.2969854.
- [13] M. Bordoloi and S. K. Biswas, “Sentiment analysis: A survey on design framework, applications and future scopes,” *Artif. Intell. Rev.*, vol. 56, no. 11, pp. 12505–12560, 2023, doi: 10.1007/s10462-023-10442-2.
- [14] R. Kusumaningrum, I. Z. Nisa, R. Jayanto, R. P. Nawangsari, and A. Wibowo, “Deep learning-based application for multilevel sentiment analysis of Indonesian hotel reviews,” *Heliyon*, vol. 9, no. 6, 2023, doi: 10.1016/j.heliyon.2023.e17147.
- [15] A. P. Giovani, A. Ardiansyah, T. Haryanti, L. Kurniawati, and W. Gata, “Analisis Sentimen Aplikasi Ruang Guru Di Twitter Menggunakan Algoritma Klasifikasi,” *J. Teknoinfo*, vol. 14, no. 2, p. 115, 2020, doi: 10.33365/jti.v14i2.679.
- [16] Lady Agustine Fitrana, S. Linawati, N. Herlinawati, R. Sa’adah, and S. Seimahuria, “Analisis Sentimen Pengguna Twitter Terhadap Brand Indosat Menggunakan Metode Naïve Bayes Classifier,” *JATI (Jurnal Mhs. Tek. Inform.)*, vol. 8, no. 3, pp. 4291–4297, 2024, doi: 10.36040/jati.v8i3.9866.
- [17] Khusnul Mulya Kautsar, Meike Rachmawati, and Harvi Puspa Wardani, “Pap Smear sebagai Metode Deteksi Dini Kanker Serviks,” *J. Ris. Kedokt.*, pp. 7–12, 2023, doi: 10.29313/jrk.vi.1775.
- [18] U. Khabibah, K. Adyani, and A. Rahmawati, “Faktor Risiko Kanker Serviks,” *Faletehan Heal. J.*, vol. 9, no. 3, pp. 270–277, 2022.
- [19] A. A. N. Mostafa and H. E. A. Mahmoud, “Review of Data Mining Concept and its Techniques,” *Int. J. Acad. Res. Bus. Soc. Sci.*, vol. 12, no. 6, 2022, doi: 10.6007/ijarbss/v12-i6/13135.
- [20] J. L. Cremer and G. Strbac, “A machine-learning based probabilistic perspective on dynamic security assessment,” *Int. J. Electr. Power Energy Syst.*, vol. 128, 2021, doi: 10.1016/j.ijepes.2020.106571.
- [21] S. Kumari, D. Kumar, and M. Mittal, “An ensemble approach for classification and prediction of diabetes mellitus using soft voting classifier,” *Int. J. Cogn. Comput. Eng.*, vol. 2, pp. 40–46, 2021, doi: 10.1016/j.ijcce.2021.01.001.