

Analisis Sentimen Publik terhadap Virus HMPV Berdasarkan Media Sosial X dengan Algoritma *Logistic Regression*

Feri Aldi Wijaya*, Parjito

¹Fakultas Teknik dan Ilmu Komputer, Sistem Informasi, Universitas Teknokrat Indonesia, Bandar Lampung, Indonesia

Email: ^{1,*}feri_aldi_wijaya@teknokrat.ac.id, ²djito@teknokrat.ac.id

Email Penulis Korespondensi: feri_aldi_wijaya@teknokrat.ac.id

Submitted: 23/02/2025; Accepted: 05/03/2025; Published: 07/03/2025

Abstrak—*Human Metapneumovirus* (HMPV) adalah virus yang menyerang saluran pernapasan yang menimbulkan gejala mirip flu, seperti batuk, demam, hingga hidung tersumbat. Virus ini pertama kali ditemukan pada tahun 2001, umumnya menimbulkan infeksi ringan, akan tetapi pada kelompok tertentu seperti anak-anak, lansia, orang dengan daya tahan tubuh lemah akan beresiko tinggi seperti bronkitis atau pneumonia. Berdasarkan masalah tersebut akan dilakukan analisis sentimen tanggapan masyarakat terhadap kasus virus *Human Metapneumovirus* (HMPV) menggunakan data yang diperoleh dari media X sebanyak 10.199 tweet, data dikumpulkan dalam periode 1 Desember 2024 hingga 30 Januari 2025 menggunakan Tweet Harvest di Google Colab dengan API Twitter. Penelitian ini mengaplikasikan Synthetic Minority Oversampling Technique (SMOTE) untuk mengatasi ketidakseimbangan data, dengan perbandingan data 80% untuk pelatihan dan 20% untuk pengujian. Hasil penelitian menunjukkan bahwa sebelum penerapan SMOTE algoritma regresi logistik memiliki akurasi sebesar 83%, dengan precision sentimen positif 90%, netral 80%, negatif 85%, serta recall untuk sentimen positif 09%, netral 89%, negatif 92%. Setelah penerapan SMOTE, akurasi meningkat mencapai 90%, perubahan paling signifikan terjadi pada sentimen positif dengan precision sentimen positif mencapai 90%, netral 87%, negatif 95%, serta recall untuk sentimen positif 96%, netral 90%, negatif 84%. Penelitian ini memberikan pemahaman tentang algoritma regresi logistik dalam analisis sentimen terhadap HMPV dan menjadi referensi bagi pemerintah serta organisasi kesehatan dalam merancang strategi komunikasi dan intervensi yang lebih efektif.

Kata Kunci: Human Metapneumovirus; Analisis Sentimen; Media Sosial; Regresi Logistik; SMOTE.

Abstract—*Human Metapneumovirus* (HMPV) is a virus that affects the respiratory tract, causing flu-like symptoms such as cough, fever, and nasal congestion. This virus was first discovered in 2001 and generally causes mild infections. However, certain groups, such as children, the elderly, and individuals with weakened immune systems, are at higher risk of developing severe conditions like bronchitis or pneumonia. Based on this issue, a sentiment analysis of public responses to Human Metapneumovirus (HMPV) cases was conducted using data collected from the X platform, consisting of 10,199 tweets. The data was gathered between December 1, 2024, and January 30, 2025, using Tweet Harvest in Google Colab with the Twitter API. This study applied the Synthetic Minority Oversampling Technique (SMOTE) to address data imbalance, with an 80% to 20% split between training and testing data. The results showed that before applying SMOTE, the logistic regression algorithm had an accuracy of 83%, with precision for positive sentiment at 90%, neutral at 80%, negative at 85%, while recall for positive sentiment was 89%, neutral 89%, negative 92%. After applying SMOTE, accuracy increased to 90%, with the most significant improvement observed in positive sentiment. The precision for positive sentiment reached 90%, neutral 87%, and negative 95%, while recall for positive sentiment was 96%, neutral 90%, negative 84%. This research provides insights into the use of logistic regression algorithms in sentiment analysis related to HMPV and serves as a reference for governments and health organizations in designing more effective communication strategies and interventions.

Keywords: Human Metapneumovirus; Sentiment Analysis; Social Media; Logistic Regression; SMOTE

1. PENDAHULUAN

Human Metapneumovirus (HMPV) adalah virus yang menyerang saluran pernapasan dan dapat menimbulkan gejala yang menyerupai flu, seperti batuk, demam, serta hidung tersumbat. Virus ini pertama kali ditemukan pada tahun 2001 oleh para ilmuwan yang meneliti infeksi pernapasan pada manusia [1][2]. Secara umum, infeksi HMPV bersifat ringan, tetapi pada kelompok dengan risiko tinggi, seperti anak-anak, lansia, dan individu dengan sistem kekebalan tubuh yang lemah, infeksi ini dapat berkembang menjadi kondisi yang lebih serius, seperti *bronkitis* atau *pneumonia*, yang berpotensi memerlukan perawatan medis lebih lanjut [3]. Penyebaran HMPV telah tercatat di berbagai belahan dunia, dengan dampak yang signifikan terhadap kesehatan masyarakat, terutama pada musim tertentu di mana kasus infeksi saluran pernapasan meningkat [4].

Permasalahan utama dalam penelitian ini adalah kurangnya pemahaman publik terhadap penyebaran dan dampak HMPV, serta bagaimana persepsi masyarakat terhadap virus ini berkembang di era digital. Media sosial telah menjadi sumber utama bagi masyarakat dalam memperoleh informasi terkait kesehatan, namun banyaknya informasi yang tersebar dapat memicu disinformasi yang berpotensi menimbulkan kepanikan atau kelalaian dalam pencegahan penyakit. Oleh karena itu, diperlukan analisis mendalam mengenai bagaimana opini publik terhadap HMPV berkembang, terutama di platform media sosial. Seiring dengan kemajuan teknologi dan meningkatnya penggunaan internet, media sosial telah menjadi platform utama bagi masyarakat untuk berbagi informasi serta opini mengenai berbagai isu kesehatan. Salah satu platform media sosial yang banyak digunakan adalah X (sebelumnya dikenal sebagai Twitter). Hingga Oktober 2023, X memiliki sekitar 666,2 juta pengguna aktif di seluruh dunia, menjadikannya platform media sosial dengan jumlah pengguna terbesar ke-12 secara global. Di Indonesia, jumlah pengguna X mencapai 27,5 juta, menempatkan Indonesia di peringkat keempat dunia setelah Amerika Serikat, Jepang, dan India [5]. Platform X memungkinkan pengguna untuk mendiskusikan berbagai topik secara real-time, menyebarkan berita,

serta memberikan respons terhadap isu kesehatan seperti HMPV, baik dalam bentuk diskusi publik maupun percakapan pribadi. Dengan sifatnya yang interaktif, X menjadi sumber data yang sangat berharga untuk memahami pola penyebaran informasi, opini, serta kekhawatiran masyarakat terhadap suatu penyakit menular [6].

Solusi yang diharapkan dalam penelitian ini adalah menggunakan analisis sentimen berbasis kecerdasan buatan untuk mengidentifikasi pola opini masyarakat terhadap HMPV di media sosial. Dengan memahami sentimen publik, diharapkan dapat diperoleh wawasan mengenai bagaimana persepsi masyarakat terbentuk, sehingga lembaga kesehatan dapat menyusun strategi komunikasi yang lebih efektif dalam memberikan edukasi mengenai HMPV dan langkah-langkah pencegahannya. Analisis ini juga dapat membantu dalam mengidentifikasi tren perbincangan terkait penyakit ini serta mengukur dampak kampanye kesehatan yang telah dilakukan.

Untuk memahami sikap publik terhadap HMPV, penelitian ini menggunakan metode analisis sentimen. Analisis sentimen adalah proses ekstraksi dan identifikasi opini atau perasaan dari suatu teks, yang dapat dikategorikan ke dalam beberapa tingkat, seperti "*Extremely Positive*" (sangat positif), "*Positive*" (positif), "*Neutral*" (netral), "*Negative*" (negatif), dan "*Extremely Negative*" (sangat negatif) [7]. Metode ini memungkinkan peneliti untuk menilai bagaimana suatu isu diterima oleh masyarakat serta bagaimana opini publik berkembang seiring waktu. Salah satu metode yang umum digunakan dalam analisis sentimen adalah regresi logistik [8]. Regresi logistik (*Logistic Regression*) merupakan algoritma pembelajaran mesin yang digunakan untuk tugas klasifikasi dengan cara memprediksi probabilitas suatu kategori berdasarkan variabel independen. Algoritma ini menggunakan fungsi logit (sigmoid) yang memetakan keluaran ke dalam rentang probabilitas antara 0 dan 1, sehingga memungkinkan klasifikasi opini berdasarkan probabilitas tertentu. *Logistic Regression* sering diterapkan dalam analisis sentimen serta berbagai aplikasi lain yang membutuhkan prediksi hasil kategorikal, seperti diagnosis penyakit atau deteksi penipuan.

Keunggulan utama dari metode ini adalah kesederhanaannya, kemudahan interpretasi hasil, serta kinerjanya yang baik, terutama pada dataset yang memiliki hubungan linier antara variabel independen dan probabilitas keluaran [9]. Selain itu, metode ini dapat dikombinasikan dengan teknik pemrosesan bahasa alami (NLP) untuk meningkatkan akurasi analisis sentimen dengan mempertimbangkan konteks kata dalam teks yang dianalisis. Dengan menggunakan metode ini, penelitian dapat memperoleh wawasan yang lebih mendalam mengenai bagaimana opini publik terbentuk dan tersebar melalui media sosial terkait isu HMPV. Informasi ini dapat digunakan oleh pemerintah dan organisasi kesehatan untuk menyusun strategi komunikasi yang lebih efektif dalam menangani penyebaran informasi dan mitigasi dampak penyakit ini.

Berbagai penelitian telah dilakukan untuk memahami analisis sentimen terhadap isu kesehatan di media sosial, khususnya terkait penyakit menular. Studi-studi terdahulu menyoroti metode dan dampaknya terhadap persepsi masyarakat. Misalnya, Marizal & Nurmita (2021) menggunakan regresi logistik ordinal untuk menganalisis persepsi mahasiswa terhadap vaksinasi COVID-19 dengan akurasi 64% [10]. Lidinillah et al. (2023) membandingkan *Logistic Regression* dan SVM dalam analisis sentimen di platform Steam, di mana SVM unggul dengan akurasi 91,15% [11]. Fazrin et al. (2023) menunjukkan bahwa *Logistic Regression* dengan *VaderSentiment* mencapai AUC 0,9254 dalam analisis sentimen vaksinasi COVID-19 di Twitter [12]. Kelvin et al. (2022) juga membandingkan *Logistic Regression* dan SVM dalam sentimen COVID-19 di Twitter dengan hasil serupa, SVM lebih unggul [13]. Sementara itu, Prabowo et al. (2023) menemukan bahwa regresi logistik multinomial menunjukkan dominasi sentimen *negatif* terhadap COVID-19 dengan akurasi 64%. [14].

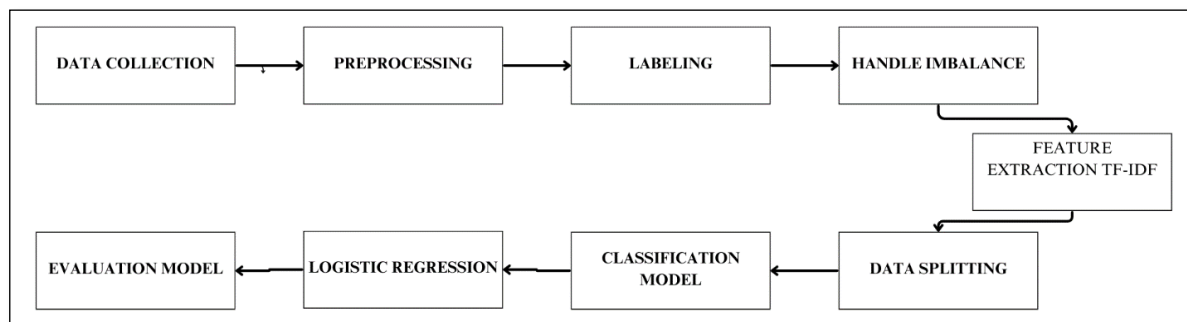
Berdasarkan penelitian sebelumnya, terdapat beberapa kesenjangan yang belum terjawab. Sebagian besar penelitian masih berfokus pada analisis sentimen terhadap COVID-19 dan vaksinasi, sementara studi tentang *Human Metapneumovirus* (HMPV) masih sangat terbatas. Selain itu, penelitian yang menggunakan regresi logistik umumnya tidak membahas pengaruh teknik penyeimbangan data seperti Synthetic Minority Oversampling Technique (SMOTE) dalam meningkatkan akurasi model. Banyak penelitian lebih menekankan perbandingan algoritma tanpa mempertimbangkan ketidakseimbangan data, padahal aspek ini menjadi tantangan utama dalam analisis sentimen. Selain itu, belum ada studi yang secara spesifik mengeksplorasi bagaimana sentimen masyarakat terhadap HMPV dapat membantu pengambilan kebijakan kesehatan atau meningkatkan efektivitas penyebaran informasi medis.

Penelitian ini bertujuan untuk menganalisis sentimen masyarakat terhadap HMPV menggunakan algoritma regresi logistik serta mengevaluasi dampak penggunaan SMOTE dalam meningkatkan performa model. Diharapkan hasil penelitian ini dapat memberikan wawasan lebih mendalam mengenai opini publik terhadap HMPV, membantu pengambil kebijakan dalam merancang strategi komunikasi kesehatan yang lebih efektif, serta berkontribusi dalam pengembangan metode analisis sentimen yang lebih akurat dan relevan.

2. METODOLOGI PENELITIAN

2.1 Tahapan Penelitian

Penelitian ini dilakukan dengan mengikuti beberapa tahapan yang sistematis untuk memastikan hasil yang valid dan dapat dipertanggungjawabkan. Setiap langkah penelitian dirancang untuk mendalami topik secara menyeluruh dan memberikan pemahaman yang lebih baik mengenai permasalahan yang diangkat. Keseluruhan tahapan penelitian ini divisualisasikan dalam bentuk skema pada Gambar 1.



Gambar 1. Skema Penelitian

Berdasarkan Gambar 1 tahap penelitian dimulai dengan proses pengumpulan data adalah awalan dari penelitian ini, dengan menggunakan metode *crawling* melalui *library Tweet Harvest* di *Google Colab* dengan memanfaatkan API Token Twitter [15]. Data berupa *tweet* dengan kata kunci tertentu dikumpulkan dalam rentang waktu yang telah ditentukan, selanjutnya data disimpan dalam format CSV (*Comma Separated Values*) untuk diproses lebih lanjut. Tahap berikutnya adalah pra-pemrosesan data elemen yang tidak relevan, akan di bersihkan dan dihilangkan teksnya. Setelah itu, data diberi label sesuai kategori tertentu, seperti sentimen positif, netral dan negatif. Untuk mengatasi ketidakseimbangan data, dilakukan teknik *Synthetic Minority Oversampling Technique* (SMOTE) agar distribusi kelas lebih merata. Selanjutnya, fitur teks diekstraksi menggunakan *Term Frequency-Inverse Document Frequency* (TF-IDF) agar menghasilkan gambaran dalam bentuk angka. Dataset kemudian dipisahkan antara data latih dan data uji untuk membangun model klasifikasi dengan algoritma seperti *Logistic Regression*. Terakhir, performa model dievaluasi menggunakan metrik seperti *accuracy*, *precision*, *recall*, dan *F1-score* untuk mengevaluasi kinerja model dalam pengklasifikasian data yang tepat.

2.2 Data Collection

Proses pengumpulan data menggunakan metode *crawling*, yaitu pengambilan informasi dari situs web secara otomatis dengan bantuan perangkat lunak [16]. Data dikumpulkan menggunakan Teknik *crawling* melalui *library tweet-harvest* dari *Node.js* [17]. Salah satu kelebihan *Tweet-Harvest* adalah kemampuannya untuk mengumpulkan data dalam jumlah besar. Proses ini dapat dijalankan melalui *Command Line Interface* (CLI) dengan hanya memanfaatkan *auth_token* [18]. Data dikumpulkan dalam tweet berbahasa Indonesia dengan kata kunci "*Human Metapneumovirus*". Sebanyak 10.199 tweet berhasil diperoleh dalam rentang waktu 1 Desember 2024 hingga 30 Januari 2025.

2.3 Data Preprocessing

Preprocessing data merupakan langkah awal dalam pengolahan data yang bertujuan untuk mempermudah analisis dan pemrosesan lebih lanjut. Proses ini mencakup pembersihan data dari gangguan, penyederhanaan dimensi data, serta penyusunan data agar lebih terstruktur [19]. Tahap preprocessing penting untuk meningkatkan kualitas hasil analisis. Proses ini mencakup pembersihan data, normalisasi teks melalui *case folding*, pemecahan teks menjadi token (*tokenizing*), penghapusan kata tidak penting (*stopwords removal*), serta reduksi kata ke bentuk dasar (*stemming*) [20]. Berikut Penjelasannya :

- Pembersihan Data (*Data Cleaning*) dilakukan untuk menghapus elemen yang tidak relevan dalam teks, seperti tanda baca, angka, URL, dan karakter khusus, sehingga data menjadi lebih bersih dan siap untuk diproses lebih lanjut. Sebagai contoh, teks "Halo! Selamat datang di #Event2024. Info lengkap: www.event.com" setelah melalui proses pembersihan akan menjadi "Halo Selamat datang di Event".
- Normalisasi Teks (*Case Folding*) bertujuan untuk mengubah seluruh huruf dalam teks menjadi huruf kecil guna menjaga konsistensi dan menghindari perbedaan akibat penggunaan huruf kapital. Sebagai contoh, teks "Baik" dan "baik" tanpa normalisasi dianggap berbeda, tetapi setelah normalisasi keduanya menjadi "baik".
- Tokenisasi (*Tokenizing*) merupakan proses pemecahan teks menjadi unit-unit kecil atau token agar dapat diolah lebih mudah dalam analisis lebih lanjut. Sebagai contoh, teks "Saya suka belajar data science" setelah tokenisasi akan menjadi ["Saya", "suka", "belajar", "data", "science"].
- Normalisasi (*normalized*) proses standarisasi kata dengan mengubah huruf menjadi kecil, mengganti kata tidak baku ke bentuk standar, serta menghapus karakter khusus, misalnya teks "Aku lg di stasiun, nunggu kereta. Gpp sih telat dikit, yg pntg aman :)" setelah normalisasi menjadi "aku lagi di stasiun menunggu kereta tidak apa-apa sih telat sedikit yang penting aman".
- Penghapusan Kata Tidak Penting (*Stopwords Removal*) dilakukan dengan menghilangkan kata-kata umum yang tidak memiliki makna signifikan dalam analisis, seperti dan, di, ke, dan yang. Sebagai contoh, teks "Saya sedang belajar di universitas" setelah melalui proses *stopwords removal* akan menjadi ["belajar", "universitas"].
- Stemming* bertujuan untuk mereduksi kata ke bentuk dasarnya dengan menghilangkan imbuhan (prefix, suffix, infix), sehingga kata-kata yang memiliki arti serupa dapat dikenali sebagai satu entitas yang sama. Sebagai contoh, kata "berlari", "berjalan", dan "menyanyikan" setelah *stemming* akan menjadi ["lari", "jalan", "nyanyi"].

2.4 Handle Imbalance

Proses pelabelan menghasilkan data yang tidak seimbang, sehingga penanganan lebih lanjut dibutuhkan menggunakan metode SMOTE (*Synthetic Minority Over-sampling Technique*). SMOTE adalah teknik pembelajaran mesin yang dirancang untuk mengatasi masalah ketidakseimbangan kelas dalam dataset. Teknik ini bekerja dengan membuat sampel baru dari kelas minoritas untuk menyeimbangkan dataset. Dengan menambahkan contoh dari kelas minoritas, dataset pelatihan menjadi lebih seimbang, dan data tambahan tersebut digunakan untuk melatih model pengklasifikasi [21].

2.5 Feature Extraction

TF-IDF adalah metode yang menggabungkan dua konsep, yaitu *term frequency* (TF) dan *inverse document frequency* (IDF). Metode ini digunakan untuk mengukur pentingnya suatu kata dalam sebuah data. TF menggambarkan seberapa sering suatu kata muncul dalam data tertentu, yang menandakan seberapa relevan kata tersebut dalam data tersebut [22]. Secara matematis algoritma ini dapat dituliskan seperti pada persamaan 1.

$$TF - IDF(t, d, D) = TF(t, d) \times IDF(t, D) \quad (1)$$

Term Frequency (TF) dan *Inverse Document Frequency* (IDF) adalah metode yang digunakan untuk mengukur pentingnya suatu kata dalam sebuah dokumen dibandingkan dengan seluruh koleksi dokumen. TF menunjukkan seberapa sering kata t muncul dalam data atau dokumen d yang sedang dianalisis. Semakin tinggi frekuensinya, semakin penting kata tersebut dalam dokumen tersebut. Sementara itu, IDF mengukur seberapa penting suatu kata dalam seluruh koleksi data D dengan mempertimbangkan seberapa banyak dokumen yang mengandung kata tersebut. Semakin jarang kata muncul dalam koleksi dokumen, semakin tinggi nilai IDF-nya, menandakan bahwa kata tersebut lebih spesifik. TF-IDF merupakan hasil perkalian antara nilai TF dan IDF, yang memberikan ukuran seberapa penting kata t dalam dokumen d dengan mempertimbangkan distribusinya dalam seluruh koleksi dokumen D .

2.6 Data Splitting

Tahap berikutnya dalam penelitian ini adalah membagi data menjadi dua bagian: data latih (*training data*) dan data uji (*testing data*), dengan pembagian sebesar 80% untuk pelatihan dan 20% untuk pengujian [23]. Data latih digunakan untuk mengembangkan model dengan menyesuaikan parameter agar mampu melakukan prediksi atau klasifikasi secara akurat. Sementara itu, data uji berfungsi untuk mengevaluasi kinerja model terhadap data yang belum pernah diproses sebelumnya. Pendekatan ini bertujuan untuk memastikan model dapat menggeneralisasi data dengan baik, menghindari *overfitting*, serta tetap efektif dalam menghadapi data baru.

2.7 Classification Model Logistic Regression

Regresi Logistik adalah sebuah metode statistik yang digunakan untuk menganalisis hubungan antara satu atau lebih variabel independen dengan variabel dependen. Metode ini bekerja dengan menghitung probabilitas terjadinya suatu peristiwa berdasarkan data yang tersedia. Probabilitas tersebut selalu berada dalam rentang antara 0 hingga 1 [24]. Secara matematis algoritma ini dapat dituliskan seperti pada persamaan 2.

$$P(y = j|X) = \frac{e^{\beta_j X}}{\sum_{k=3}^k e^{\beta_j X}} \quad (2)$$

$P(y = j|X)$ menunjukkan probabilitas bahwa input X diklasifikasikan ke dalam kategori j . Nilai eksponensial dari kombinasi linier antara fitur X dan parameter β_j menunjukkan seberapa besar kontribusi fitur terhadap kategori tertentu. Untuk memastikan total probabilitas selalu berjumlah 1, dilakukan normalisasi dengan menjumlahkan eksponensial dari semua kategori yang tersedia, yaitu $\sum_{k=3}^k e^{\beta_j X}$. Dalam kasus analisis sentimen, terdapat tiga kategori utama, yaitu positif, netral, dan negatif, sehingga nilai $k = 3$. Dengan menerapkan fungsi ini, model dapat menghitung probabilitas setiap kategori untuk suatu input X dan memilih kategori dengan probabilitas tertinggi sebagai hasil klasifikasi.

2.8 Evaluasi Model

Selanjutnya adalah melakukan evaluasi untuk menilai kinerja model klasifikasi yang digunakan. Evaluasi ini dilakukan menggunakan confusion matrix, dengan beberapa metrik utama. Akurasi (*Accuracy*) digunakan untuk mengukur sejauh mana prediksi model sesuai dengan data secara keseluruhan, baik untuk kategori positif maupun negatif. *Precision* membandingkan jumlah prediksi yang benar terhadap total prediksi positif yang dibuat oleh model. *Recall* mengukur seberapa banyak data positif yang berhasil diprediksi dengan benar dibandingkan dengan total data positif yang sebenarnya. Selain itu, F1-Score dihitung sebagai nilai rata-rata harmonis antara *Precision* dan *Recall* untuk memberikan gambaran seimbang mengenai kinerja model [25].

$$Akurasi = \frac{TP+TN}{TP+TN+FP+FN} \times 100\% \quad (3)$$

$$Presisi = \frac{TP}{TP+FP} \quad (4)$$



$$Recall = \frac{TP}{TP+FN} \tag{5}$$

$$f_1 - score = 2x = \frac{Presisi \times Recall}{Presisi + Recall} \tag{6}$$

3. HASIL DAN PEMBAHASAN

3.1 Data Collection

Data dikumpulkan melalui metode crawling dengan memanfaatkan *library Tweet Harvest* di *Google Colab* dan *API Token Twitter*. API Token digunakan untuk mengautentikasi aplikasi atau skrip yang mengakses data dari *Twitter* melalui API resmi. Hasil pengumpulan data dirangkum pada Tabel 1.

Tabel 1. Hasil Pengumpulan Data

No	Tweet
1	<i>Human Metapneumovirus</i> HMPV: A Growing Respiratory Concern! https://t.co/31Cpu8VKGH #HMPVvirus #India #NewVirus #Covid19
10.199	Banyak yang keblinger HMPV yang sekarang lagi ngetrend itu BUKAN Virus Baru. (Belajar bikin utas pendek) Kemunculan HMPV sebagai emerging infection itu sudah dideteksi oleh para ilmuwan sejak tahun 2001. Risetnya ada disini https://t.co/bBKyEzrCBy (Cont) https://t.co/kpPPr5EO1t

Berdasarkan Tabel 1 yang menunjukkan hasil pengumpulan data, yang diakumulasikan berupa tweet berbahasa Indonesia yang memiliki kata kunci "*Human Metapneumovirus*" selama periode 1 Desember 2024 hingga 30 Januari 2025. Sebanyak 10.119 tweet berhasil dikumpulkan dan disimpan dalam format CSV untuk tahap *preprocessing* selanjutnya.

3.2 Preprocessing

Dari periode Desember 2024 hingga Januari 2025, sebanyak 10.199 tweet berhasil dikumpulkan. Selanjutnya, data melaluserangkaian proses *preprocessing* untuk mempersiapkannya sebelum dianalisis. Pada tahap ini, data akan menjadi terorganisir, bersih, juga sudah bisa untuk dianalisis setelah melewati seluruh tahapan. Dapat dilihat pada Tabel 2.

Tabel 2. Hasil Preprocessing

Tahapan	Tweet
Data Tweet	Virus <i>Human Metapneumovirus</i> (HMPV) dilaporkan telah ditemukan di Indonesia. Semua kasus yg ditemukan melibatkan anak2. NOW Simak bahasannya dlm HEALTH CORNER brsm Dr.Santi @kenapayadok 19.00-20.00 YouTube: https://t.co/dlkRxp61do Streaming: https://t.co/TRcBK95osQ
Data Cleaning	Virus <i>Human Metapneumovirus</i> HMPV dilaporkan telah ditemukan di Indonesia Semua kasus yg ditemukan melibatkan anak NOW Simak bahasannya dlm HEALTH CORNER brsm Dr Santi YouTube Streaming
Case Folding	virus human metapneumovirus hmpv dilaporkan telah ditemukan di indonesia semua kasus yg ditemukan melibatkan anak now simak bahasannya dlm health corner brsm dr santi youtube streaming
Tokenizing	[virus, human, metapneumovirus, hmpv, dilaporkan, telah, ditemukan, di, indonesia, semua, kasus, yg, ditemukan, melibatkan, anak, now, simak, bahasannya, dlm, health, corner, brsm, dr, santi, youtube, streaming,]
Normalizez	[virus, human, metapneumovirus, hmpv, dilaporkan, telah, ditemukan, di, indonesia, semua, kasus, yang, ditemukan, melibatkan, anak, now, simak, bahasannya, dlm, health, corner, brsm, dari, santi, youtube, streaming,]
StopWords	[virus, human, metapneumovirus, hmpv, dilaporkan, ditemukan, indonesia, ditemukan, melibatkan, anak, now, simak, bahasannya, dlm, health, corner, brsm, santi, youtube, streaming,]
Stemming	virus human metapneumovirus hmpv lapor temu indonesia temu libat anak now simak bahas dlm health corner brsm santi youtube streaming

Berdasarkan Tabel 2 yang menunjukkan hasil Proses *preprocessing* data dilakukan untuk membersihkan dan menyiapkan teks agar dapat dianalisis dengan lebih efektif. Tahap pertama adalah Data Tweet, di mana teks masih berupa data mentah yang mengandung link, tanda baca, dan singkatan, seperti: "Virus Human Metapneumovirus (HMPV) ... <https://t.co/TRcBK95osQ>." Kemudian, dilakukan Data Cleaning dengan menghapus elemen-elemen yang tidak diperlukan, seperti link dan karakter khusus, menghasilkan teks yang lebih sederhana: "Virus Human Metapneumovirus HMPV ... YouTube Streaming." Selanjutnya, tahap Case Folding dilakukan dengan mengubah semua huruf menjadi huruf kecil untuk menjaga konsistensi dalam analisis. Teks yang telah diproses menjadi: "virus human metapneumovirus hmpv ... youtube streaming." Setelah itu, tahap Tokenizing memecah teks menjadi kata-

kata individu, sehingga lebih mudah untuk diproses lebih lanjut. Hasil dari tahap ini adalah daftar kata-kata seperti: [virus, human, ... youtube, streaming]. Pada tahap Normalization, kata-kata yang tidak baku atau memiliki singkatan diubah ke dalam bentuk standar, misalnya "yg" menjadi "yang" dan "brsm" menjadi "bersama". Selanjutnya, tahap Stop Words Removal menghapus kata-kata umum yang tidak memiliki makna signifikan dalam analisis, seperti "di", "yang", dan "telah". Hasilnya, teks menjadi lebih padat dengan hanya menyisakan kata-kata yang lebih bermakna: [virus, human, ... youtube, streaming]. Terakhir, tahap Stemming dilakukan untuk mengubah kata berimbuhan menjadi bentuk dasarnya, seperti "dilaporkan" menjadi "lapor" dan "melibatkan" menjadi "libat". Hasil akhirnya adalah teks yang lebih ringkas dan siap untuk dianalisis: "virus human metapneumovirus ... youtube streaming."

3.3 Labeling

Setelah melalui tahap pemrosesan data, sebanyak 5.647 tweet berhasil disaring menjadi data yang bersih dan siap untuk dianalisis. Selanjutnya, setiap tweet dilabeli menggunakan lexicon based berdasarkan analisis sentimen Dapat dilihat pada Tabel 3 berikut untuk hasil pelabelan:

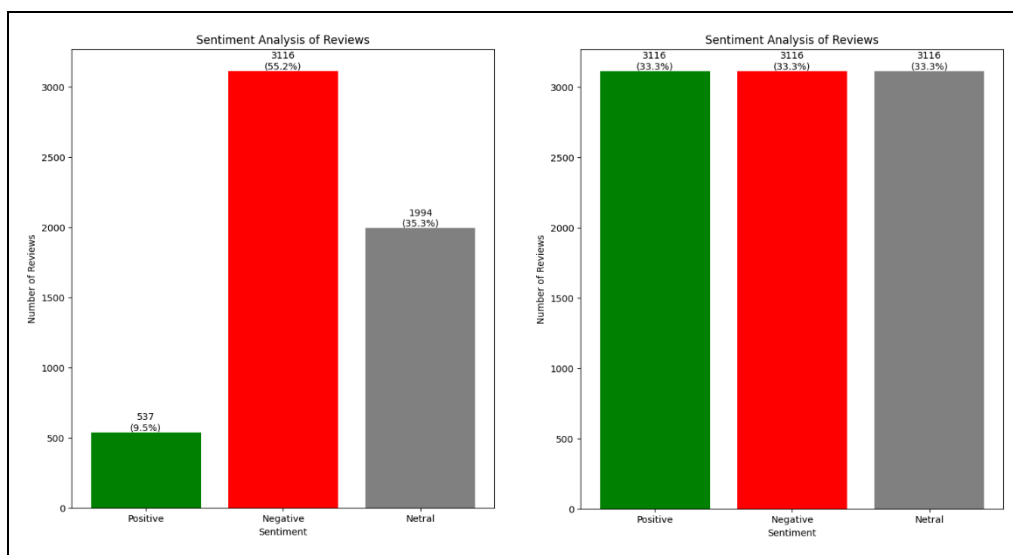
Tabel 3. Hasil Labeling

No	Tweet	Sentimen Score	Label
1	virus hmpv human metapneumovirus rebak wilayah tiengkong lapor temu indonesia temu libat anak anak menteri sehat budi gunadi sadikin masyarakat panik hmpv virus	-27	Negatif
2	virus human metapneumovirus hmpv lapor temu indonesia temu libat anak simak bahas dlm health corner brsm santi senin jan youtube sonora fm streaming	0	Netral
5.647	virus human metapneumovirus hmpv perhati tenang ya dulur stayhealthy waspadahmpv infokesehatan kominfodiy jogjaistimewa	4	Positif

Berdasarkan Tabel 3, proses pelabelan dilakukan dengan menghitung nilai polaritas dalam rentang -1 hingga 1. Nilai di atas 0 dianggap positif, sementara di bawah 0 dianggap negative sedangkan nilai 0 dianggap netral.

3.4 Handle Imbalance (SMOTE)

Hasil awal analisis menunjukkan bahwa terdapat 537 tweet positif, 3.116 tweet negative, dan 1.994 tweet netral. Yang menunjukan ketidakseimbangan data mayoritas dan minoritas, Untuk mengatasi ketidakseimbangan data ini, digunakan teknik SMOTE. Hasil perbandingan distribusi sentimen dan dilihat pada Gambar 2.



Gambar 2. Perbandingan sentiment penggunaan Teknik SMOTE

Gambar 2 menunjukkan perbandingan sebelum dan setelah penerapan SMOTE. Sebelum penerapan metode SMOTE, diagram batang memperlihatkan perbedaan signifikan dalam distribusi sentimen, dengan 537 *tweet* positif, 3.116 *tweet* negative, dan 1.994 *tweet* netral. Ketidakseimbangan ini dapat menyebabkan bias pada model klasifikasi, di mana model lebih cenderung memprediksi kelas mayoritas, dengan akurasi yang lebih tinggi, namun kurang efektif dalam memprediksi kelas minoritas. Setelah penerapan SMOTE, jumlah *tweet* dalam kedua kategori menjadi seimbang, masing-masing 3.166. Penyeimbangan ini penting untuk meningkatkan akurasi analisis sentimen secara keseluruhan, karena model dapat belajar secara adil dari kedua kelas, mengurangi bias, dan meningkatkan kinerja prediksi pada data minoritas. Penyesuaian ini memastikan bahwa model klasifikasi memberikan hasil yang lebih andal dan representatif.

3.5 Hasil Feature Extraction (TF-IDF)

TF-IDF (*Term Frequency - Inverse Document Frequency*) merupakan metode untuk menilai kepentingan suatu kata dalam sebuah dokumen relatif terhadap kumpulan dokumen lainnya. TF mengukur seberapa sering kata muncul dalam dokumen, sedangkan IDF menunjukkan seberapa jarang kata tersebut muncul di seluruh dokumen, sehingga kata-kata umum memiliki bobot lebih rendah. Teknik ini sering digunakan dalam analisis teks, seperti pencarian informasi dan pemrosesan bahasa alami (NLP).

Tabel 4. Hasil TF-IDF

Term	TF	IDF	TF-IDF
di	0.13	2.03	0.25
hmpv	0.13	1.09	0.14
human	0.13	1.77	0.22
metapneumovirus	0.13	1.74	0.22
pada	0.13	3.36	0.42
pencegahan	0.13	4.29	0.54
penyebaran	0.13	4.57	0.57
senin	0.13	6.03	0.75

Tabel 4 menampilkan hasil perhitungan TF-IDF (*Term Frequency - Inverse Document Frequency*) untuk sebuah teks terkait rapat koordinasi pencegahan penyebaran *Human Metapneumovirus* (HMPV). Nilai TF menunjukkan frekuensi relatif suatu kata dalam dokumen, sementara IDF mengukur kepentingan kata berdasarkan kemunculannya di seluruh dokumen. Kata dengan nilai TF-IDF tertinggi, seperti "senin" (0.753), "penyebaran" (0.571), dan "pencegahan" (0.536), dianggap lebih signifikan dalam dokumen ini dibandingkan kata umum seperti "di" (0.254) dan "hmpv" (0.136).

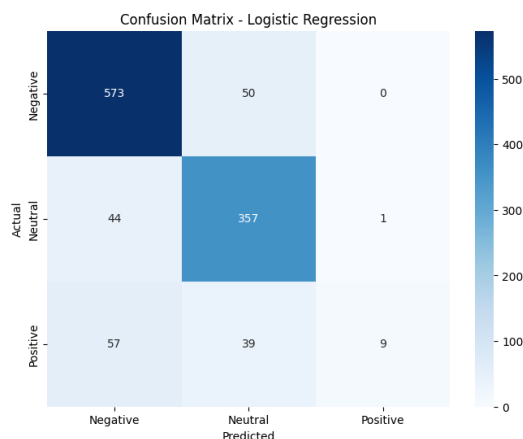
3.6 Hasil Klasifikasi Model

Algoritma *Logistic Regression* menggunakan perbandingan data training sebesar 80% dan data testing 20%. Evaluasi ini dilaksanakan dengan 2 kondisi, yaitu sebelum dan setelah penerapan SMOTE. Proses evaluasi tersebut dilaksanakan pada data yang telah dikumpulkan dan siap untuk dianalisis. Tabel 4 berikut ini menunjukkan skor *precision*, *accuracy*, *recall*, dan *FI*

Tabel 4. Perbandingan Klasifikasi Sebelum dan Setelah SMOTE

Matriks	Sebelum SMOTE	Setelah SMOTE
<i>Accuracy</i>	83%	90%
Sentimen Negatif		
<i>Precision</i>	85%	95%
<i>Recall</i>	92%	84%
<i>F1-Score</i>	88%	89%
Sentimen Netral		
<i>Precision</i>	80%	87%
<i>Recall</i>	89%	90%
<i>F1-Score</i>	84%	88%
Sentimen Positif		
<i>Precision</i>	90%	90%
<i>Recall</i>	09%	96%
<i>F1-Score</i>	16%	93%

Hasil pengujian menunjukkan bahwa model tersebut menghasilkan performa yang memuaskan, dengan akurasi meningkat dari 83% menjadi 90%. Pada sentimen negatif, *precision* naik dari 85% menjadi 95%, menunjukkan klasifikasi yang lebih akurat, meskipun *recall* menurun dari 92% menjadi 84%, dan *F1-score* bertambah dari 88% menjadi 89%. Untuk sentimen netral, *precision* meningkat dari 80% menjadi 87%, *recall* naik dari 89% menjadi 90%, dan *F1-score* bertambah dari 84% menjadi 88%, menunjukkan perbaikan keseimbangan antara *precision* dan *recall*. Perubahan paling signifikan terjadi pada sentimen positif, 90% untuk *precision* negative maupun positif, di mana *recall* melonjak dari 9% menjadi 96%, dan *F1-score* meningkat drastis dari 16% menjadi 93%, menandakan model jauh lebih baik dalam mengenali data positif yang sebenarnya. Dapat dilihat pada Gambar 3 dan Gambar 4.



Gambar 3. Confusion matrix sebelum SMOTE

Pada Gambar 3 terlihat bahwa analisis confusion matrix memberikan wawasan lebih mendalam tentang kemampuan klasifikasi model Logistic Regression. Model menunjukkan kecenderungan yang kuat dalam mengidentifikasi sentimen Negative, hal ini terlihat sebelum optimalisasi SMOTE dilakukan. Logistic Regression mencatat True Negative sebanyak 573, False Negative 57, True Neutral 357, False Neutral 44, sementara True Positive 9 dan False Positive 50. Model ini menunjukkan kesulitan dalam mengklasifikasikan sentimen Positive, yang kemungkinan disebabkan oleh ketidakseimbangan data. Oleh karena itu, penerapan SMOTE diharapkan dapat meningkatkan performa klasifikasi terutama pada kelas Positive dan membantu model mengenali kelas Neutral dengan lebih baik

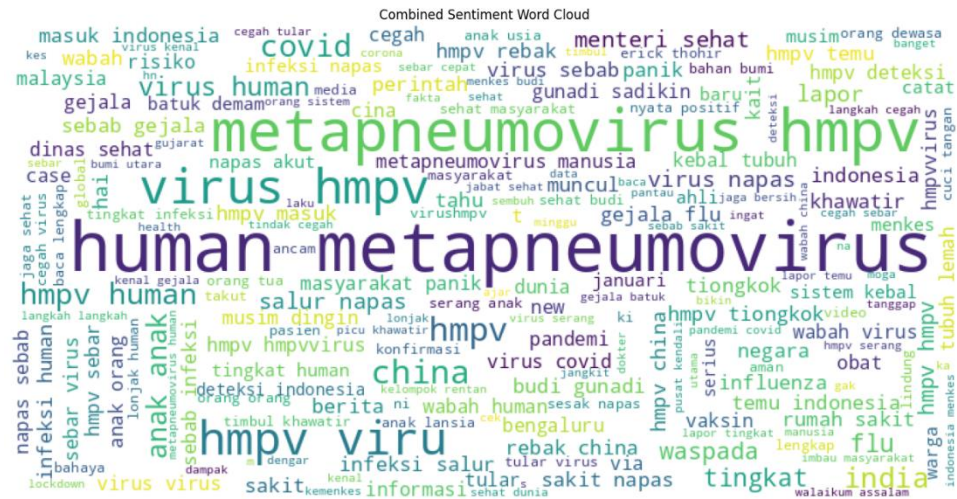


Gambar 4. Confusion matrix sesudah SMOTE

Sesudah penerapan SMOTE yang ditunjukkan pada gambar di atas, terjadi peningkatan substansial dalam kemampuan klasifikasi model Logistic Regression. Model menunjukkan peningkatan performa dengan True Negative sebanyak 539 dan False Negative 7, serta True Positive 611 dan False Positive 35. Model juga mengalami perbaikan dalam mengenali kelas Neutral, dengan True Neutral 538 dan False Neutral 23. Hasil ini mengindikasikan bahwa optimasi SMOTE berhasil. meningkatkan keseimbangan dan akurasi klasifikasi model, terutama dalam mengenali kelas Positive, yang sebelumnya sulit diklasifikasikan. Hasil optimasi SMOTE terbukti efektif dalam meningkatkan kemampuan model untuk menyeimbangkan prediksi antara kelas Negative, Neutral, dan Positive, dengan peningkatan signifikan pada True Positive, yang menunjukkan perbaikan yang lebih baik dibandingkan sebelum SMOTE diterapkan.

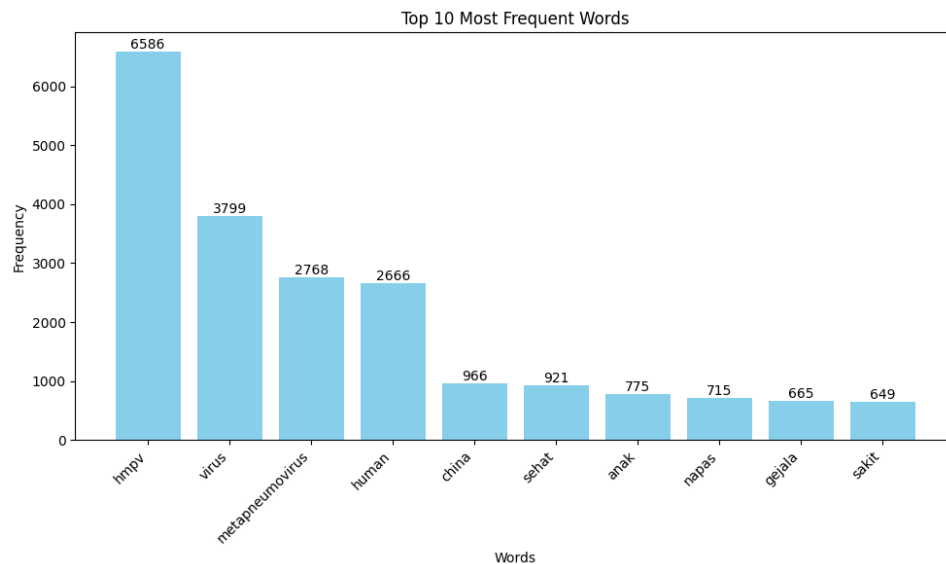
3.6 Hasil Visualisasi WordCloud

Dalam eksperimen yang telah dilakukan, visualisasi wordcloud juga digunakan untuk menganalisis data teks. Teks ditampilkan dalam bentuk cloud kata, Dengan istilah yang paling sering muncul ditampilkan dengan ukuran visual yang lebih besar, sementara kata-kata yang jarang muncul tampil lebih kecil. Kata-kata seperti "human metapneumovirus" dan "hmpv", istilah yang paling sering muncul dalam teks akan ditampilkan dengan ukuran dan warna yang disesuaikan dengan frekuensi kemunculannya.



Gambar 5. WordCloud Dengan StopWords

Visualisasi pada Gambar 5 di atas menyajikan perolehan *word cloud* yang merupakan semua sentimen. Visualisasi ini menggambarkan istilah-istilah yang paling sering digunakan oleh pengguna, memberikan gambaran tentang topik atau isu yang dominan dalam komentar seperti yang paling banyak "human metapneumovirus", "hmpv", "virus", "china", "gejala", dan "infeksi". Dengan demikian, dapat diasumsikan dalam penelitian ini bahwa terdapat perhatian tinggi dari pengguna terhadap isu kesehatan terkait human metapneumovirus (HMPV). Kata-kata seperti "gejala", "infeksi", "salur napas", dan "sakit" menunjukkan adanya diskusi mengenai dampak kesehatan virus ini, sedangkan istilah seperti "china", "indonesia", dan "dunia" mengindikasikan bahwa penyebaran virus ini menjadi perhatian global. Selain itu, adanya kata-kata seperti "waspada", "panik", dan "kebal tubuh" menunjukkan respons masyarakat terhadap ancaman kesehatan ini.



Gambar 6. Hasil grafik frekuensi kata

Gambar 6 menampilkan diagram batang yang menunjukkan 10 kata yang paling sering muncul dalam sebuah kumpulan teks. Dari diagram, kata "hmpv" memiliki frekuensi tertinggi, yaitu 6.586 kali, diikuti oleh kata "virus" dengan 3.799 kali kemunculan. Kata-kata lainnya seperti "metapneumovirus", "human", dan "china" juga memiliki jumlah kemunculan yang cukup signifikan. Sementara itu, kata-kata seperti "sehat", "anak", "napas", "gejala", dan "sakit" memiliki frekuensi yang lebih rendah, tetapi tetap masuk dalam 10 besar kata yang paling sering muncul.

4. KESIMPULAN

Berdasarkan hasil penelitian dengan menerapkan algoritma *Logistic Regression* untuk analisis sentimen masyarakat terhadap *Human Metapneumovirus* (HMPV) yang menggunakan data dari media sosial X diperoleh ketidakseimbangan data antara sentimen mayoritas dan minoritas, model cenderung lebih akurat dalam

mengklasifikasikan kelas mayoritas. Dengan menerapkan metode SMOTE diperoleh hasil kerja model lebih bagus. Hasil penelitian menunjukkan bahwa sebelum penerapan SMOTE algoritma regresi logistik memiliki akurasi sebesar 83%, dengan *precision* sentimen positif 90%, netral 80%, negatif 85%, serta recall untuk sentimen positif 09%, netral 89%, negatif 92%. Setelah penerapan SMOTE, akurasi meningkat mencapai 90%, perubahan paling signifikan terjadi pada sentimen positif dengan *precision* sentimen positif mencapai 90%, netral 87%, negatif, 95% serta recall untuk sentimen positif 96%, netral 90%, negatif 84%. Pada penelitian ini memiliki beberapa keterbatasan, seperti penggunaan data yang hanya berasal dari media sosial X, sehingga mungkin tidak sepenuhnya mewakili opini masyarakat secara umum. Penelitian selanjutnya dapat memperluas sumber data, menggunakan algoritma lain, serta mempertimbangkan metode tambahan. Meskipun SMOTE telah diterapkan, pendekatan metode lain dapat mengatasi ketidakseimbangan data bisa dipertimbangkan. Terakhir, analisis lebih lanjut terhadap faktor sosial dan kesehatan yang mempengaruhi sentimen publik terhadap HMPV dapat memberikan wawasan yang lebih mendalam.

REFERENCES

- [1] S. Ma *Et Al.*, “Development Of A Novel Multi-Epitope Mrna Vaccine Candidate To Combat Hmpv Virus,” *Hum Vaccin Immunother.*, Vol. 19, No. 3, 2023, Doi: 10.1080/21645515.2023.2293300.
- [2] S. M. Aji Muhawarman, “Wabah Virus Hmpv Merebak Di China, Kemenkes Imbau Publik Untuk Waspada,” <https://sehatnegeriku.kemkes.go.id/>. Accessed: Feb. 23, 2025. [Online]. Available: <https://sehatnegeriku.kemkes.go.id/baca/umum/20250103/5547197/wabah-virus-hmpv-merebak-di-china-kemenkes-imbau-publik-untuk-waspada/>
- [3] J. C. Muñoz-Escalante, G. Mata-Moreno, G. Rivera-Alfaro, And D. E. Noyola, “Global Extension And Predominance Of Human Metapneumovirus A2 Genotype With Partial G Gene Duplication,” *Viruses*, Vol. 14, No. 5, May 2022, Doi: 10.3390/V14051058.
- [4] Y. Feng, T. He, B. Zhang, H. Yuan, And Y. Zhou, “Epidemiology And Diagnosis Technologies Of Human Metapneumovirus In China: A Mini Review,” *Biomed Central Ltd.*, Access Date Dec. 01, 2024, Doi: 10.1186/S12985-024-02327-9.
- [5] <https://databoks.katadata.co.id/>, “Ada 27 Juta Pengguna Twitter Di Indonesia, Terbanyak Ke-4 Global,” <https://databoks.katadata.co.id/>. Accessed: Feb. 07, 2025. [Online]. Available: <https://databoks.katadata.co.id/teknologi-telekomunikasi/statistik/75dd4b36866dc54/ada-27-juta-pengguna-twitter-di-indonesia-terbanyak-ke-4-global>
- [6] N. Hadi And D. Sugiarto, “Analisis Sentimen Pembangunan Ikn Pada Media Sosial X Menggunakan Algoritma Svm, Logistic Regression Dan Naïve Bayes,” *Jurnal Informatika: Jurnal Pengembangan It*, Vol. 10, No. 1, Pp. 37–49, Jan. 2025, Doi: 10.30591/Jpit.V10i1.7106.
- [7] Kelvin, J. Banjarmasin, E. Indra, And S. H. Sinurat, “Analisis Perbandingan Sentimen Corona Virus Disease-2019 (Covid-19) Pada Twitter Menggunakan Metode Logistic Regression Dan Support Vector Machine (Svm),” *Jusikom Prima*, Vol. 5, No. 2, Feb. 2022, Accessed: Feb. 27, 2025. [Online]. Available: <https://jurnal.unprimdn.ac.id/index.php/jusikom/article/view/2365>
- [8] Y. Pratama, D. T. Murdiansyah, And K. M. Lhaksmana, “Analisis Sentimen Kendaraan Listrik Pada Media Sosial Twitter Menggunakan Algoritma Logistic Regression Dan Principal Component Analysis,” *Jurnal Media Informatika Budidarma*, Vol. 7, No. 1, P. 529, Feb. 2023, Doi: 10.30865/Mib.V7i1.5575.
- [9] Q. Philippot *Et Al.*, “Human Metapneumovirus Infection Is Associated With A Substantial Morbidity And Mortality Burden In Adult Inpatients,” *Heliyon*, Vol. 10, No. 13, Jul. 2024, Doi: 10.1016/J.Heliyon.2024.E33231.
- [10] M. Marizal And A. Nurmita, “Mendeteksi Pengaruh Persepsi Mahasiswa Terhadap Vaksinasi Covid-19 Menggunakan Ordinal Logistic Regression,” *Al-Ittizaan: Jurnal Bimbingan Konseling Islam*, Vol. 4, No. 2, P. 43, Feb. 2022, Doi: 10.24014/Ittizaan.V4i2.16201.
- [11] E. R. Lidinillah, T. Rohana, And A. R. Juwita, “Analisis Sentimen Twitter Terhadap Steam Menggunakan Algoritma Logistic Regression Dan Support Vector Machine,” *Teknosains : Jurnal Sains, Teknologi Dan Informatika*, Vol. 10, No. 2, Pp. 154–164, Jul. 2023, Doi: 10.37373/Tekno.V10i2.440.
- [12] F. Fazrin, O. N. Pratiwi, And R. Andreswari, “Perbandingan Algoritma K-Nearest Neighbor Dan Logistic Regression Pada Analisis Sentimen Terhadap Vaksinasi Covid-19 Pada Media Sosial Twitter Dengan Pelabelan Vader Dan Textblob,” *E-Proceeding Of Engineering*, Vol. 10, No. 2, P. 1596, Apr. 2023, Accessed: Feb. 27, 2025. [Online]. Available: <https://openlibrarypublications.telkomuniversity.ac.id/index.php/engineering/article/view/19960>
- [13] Kelvin, J. Banjarmasin, E. Indra, And S. H. Sinurat, “Analisis Perbandingan Sentimen Corona Virus Disease- 2019 (Covid19) Pada Twitter Menggunakan Metode Logistic Regression Dan Support Vector Machine (Svm),” *Jusikom Prima*, Vol. 5, No. 2, Pp. 2580–2879, Feb. 2022, Accessed: Feb. 27, 2025. [Online]. Available: <https://jurnal.unprimdn.ac.id/index.php/jusikom/article/view/2365>
- [14] R. Prabowo, H. Sujaini, And T. Rismawan, “Analisis Sentimen Pengguna Twitter Terhadap Kasus Covid-19 Di Indonesia Menggunakan Metode Regresi Logistik Multinomial,” *Jurnal Sistem Dan Teknologi Informasi (Justin)*, Vol. 11, No. 1, P. 85, Jan. 2023, Doi: 10.26418/Justin.V11i1.57450.
- [15] A. Shiddicky, “Analisis Sentimen Masyarakat Terhadap Kebijakan Vaksinasi Covid-19 Pada Media Sosial Twitter Menggunakan Metode Logistic Regression,” *Coscitech*, Vol. 3, No. 2, Pp. 99–106, Oct. 2022, Accessed: Feb. 27, 2025. [Online]. Available: <https://doi.org/10.37859/Coscitech.V3i2.3836>
- [16] B. R. Talisman, I. Darmawan, And I. N. Pratiwi. Oktaria, “Perancangan Aplikasi Data Crawling Untuk Pencarian Buku Pada Toko Buku Online,” *E-Proceeding Of Engineering*, Vol. 11, No. 4, 2024, Accessed: Feb. 27, 2025. [Online]. Available: <https://openlibrarypublications.telkomuniversity.ac.id/index.php/engineering/article/view/24280>
- [17] B. A. Yuniarossy *Et Al.*, “Analisis Sentimen Terhadap Isu Feminisme Di Twitter Menggunakan Model Convolutional Neural Network (Cnn),” *Jurnal Lebesgue*, Vol. 5, No. 1, Apr. 2024, Doi: 10.46306/Lb.V5i1.



- [18] R. Darman, “Analisis Sentimen Respons Twitter Terhadap Persyaratan Badan Penyelenggara Jaminan Sosial (Bpjs) Di Kantor Pertanahan,” *Widya Bhumi*, Vol. 3, No. 2, Pp. 113–136, Oct. 2023, Doi: 10.31292/Wb.V3i2.61.
- [19] Abd. A. Syam, G. Hardy M, A. Salim, D. F. Suriyanto, And M. Fajar B, “Analisis Teknik Preprocessing Pada Sentimen Masyarakat Terkait Konflik Israel-Palestina Menggunakan Support Vector Machine,” *Jipi (Jurnal Ilmiah Penelitian Dan Pembelajaran Informatika)*, Vol. 9, No. 3, Pp. 1464–1472, Aug. 2024, Doi: 10.29100/Jipi.V9i3.5527.
- [20] E. R. Lidinillah, T. Rohana, And A. R. Juwita, “Analisis Sentimen Twitter Terhadap Steam Menggunakan Algoritma Logistic Regression Dan Support Vector Machine,” *Teknosains : Jurnal Sains, Teknologi Dan Informatika*, Vol. 10, No. 2, Pp. 154–164, Jul. 2023, Doi: 10.37373/Tekno.V10i2.440.
- [21] E. G. Radjah And A. C. Talakua, “Analisis Sentimen Komentar Terhadap Konten Tenun Ntt Di Youtube Menggunakan Metode Smote Dan Logistic Regression,” *Transformatif*, Vol. Xiii, No. 2, Pp. 84–94, 2024, Accessed: Feb. 27, 2025. [Online]. Available: <https://ojs.unkriswina.ac.id/index.php/transformatif/article/view/1005>
- [22] R. Yusuf, K. Bahumatra, N. Komaria, E. A. Aqma, And L. Cahyani, “Analisis Sentimen Terhadap Aplikasi Google Meet Berdasarkan Komentar Pengguna Menggunakan Metode Logistic Regression,” *Jurnal Ilmiah Edutic : Pendidikan Dan Informatika*, Vol. 11, No. 1, Pp. 53–64, Nov. 2024, Doi: 10.21107/Edutic.V11i1.28113.
- [23] J. Kajian *Et Al.*, “Analisis Sentimen Di Twitter Terkait Tim Nasional Sepak Bola Indonesia Menggunakan Metode Support Vector Machine,” *Jurnal Kajian Dan Terapan Matematika*, Vol. 10, Pp. 96–104, Jul. 2024, Accessed: Feb. 27, 2025. [Online]. Available: <https://journal.student.uny.ac.id/jktm/article/view/19561>
- [24] M. I. Putri And I. Kharisudin, “Analisis Sentimen Pengguna Aplikasi Marketplace Tokopedia Pada Situs Google Play Menggunakan Metode Support Vector Machine (Svm), Naïve Bayes, Dan Logistic Regression,” *Prisma, Prosiding Seminar Nasional Matematika*, Vol. 5, Pp. 759–766, 2022, Accessed: Feb. 27, 2025. [Online]. Available: <https://journal.unnes.ac.id/sju/prisma/article/view/54577>
- [25] B. Setiawan, K. Ahmad Baihaqi, E. Nurlaelasari, And H. Hikmayanti Handayani, “Analisis Sentimen Ulasan Aplikasi Identitas Kependudukan Digital Menggunakan Algoritma Logistic Regression Dan K-Nearest Neighbor,” *Technology And Science (Bits)*, Vol. 6, No. 1, Pp. 533–540, 2024, Doi: 10.47065/Bits.V6i1.5389.