

Analisis Sentimen Opini Masyarakat Terhadap Virus Omicron Di Indonesia Menggunakan Metode Naïve Bayes

Aziz Mustafa¹⁾, Triana Harmini²⁾, Angga Fahri Setiawan³⁾, Nur Aini Shofiya Asy'ari⁴⁾

Program Studi Teknik Informatika, Fakultas Sains dan Teknologi, Universitas Darussalam Gontor ^{1), 2), 3)}

Program Studi Ilmu Komunikasi, Fakultas Humaniora Universitas Darussalam Gontor ⁴⁾

aziz@unida.gontor.ac.id ¹⁾, triana@unida.gontor.ac.id ²⁾, anggafahrisetiawan@unida.gontor.ac.id ³⁾,
nurainishofia@unida.gontor.ac.id ⁴⁾

Abstrak

Virus covid-19 terus bermutasi membentuk varian baru. Varian terakhir yang terdeteksi yaitu, varian Omicron dikenal sebagai varian B.1.1.529. Varian ini pertama kali dilaporkan dari Afrika Selatan pada 24 November 2021 dan saat ini telah menyebar ke seluruh dunia. Pada bulan juli 2022 kasus Omicron mengalami lonjakan. Hal ini menimbulkan banyaknya opini masyarakat khususnya di media sosial mengenai virus omicron. Penelitian ini bertujuan untuk megklasifikasi opini masyarakat terhadap kemunculan virus Omicron pada sosial media twitter dan youtube ke dalam kelas positif, negatif dan netral. Metode yang digunakan pada penelitian ini yaitu algoritma naïve bayes. Naïve bayes merupakan salah satu metode yang bisa digunakan untuk klasifikasi sentiment opini publik. Hasil penelitian sentiment analisis menggunakan naïve bayes menghasilkan tingkat akurasi sebesar 0.82%. Kemudian model diuji untuk membaca opini public di twitter dari tanggal 5 oktober 2022 sampai 27 oktober 2022. Untuk hasil sentiment pengguna twitter pada kata kunci Covid 19 didominasi oleh sentiment positif dengan presentase 85%. Dan untuk sentiment dengan kata kunci Omicron masih didominasi oleh sentiment positif dengan presentase 49%. Disebutkan dari hasil klasifikasi pada data bulan oktober 2022 berarti masyarakat jauh lebih optimis akan menghilangnya virus omicron. Untuk selanjutnya penelitian ini dapat ditingkatkan dengan menambah data atau menggunakan algoritma yang berbeda ataupun implementasi pada algoritma yang sudah ada.

Kata kunci: Covid-19, Omicron, Media Sosial, Naïve bayes

Abstract

[Analysis Of News Sentiment And Public Opinion On Omicron Virus In Indonesia Using The Naïve Bayes Method] The Covid-19 virus continues to mutate to form new variants. The last detected variant, the Omicron variant, is known as the B.1.1.529 variant. This variant was first reported from South Africa on 24 November 2021 and has now spread worldwide. In July 2022 Omicron cases experienced a spike. This has led to a lot of public opinion, especially on social media, about the omicron virus. This study aims to classify public opinion on the emergence of the Omicron virus on Twitter and YouTube social media into positive, negative and neutral classes. The method used in this study is the naïve Bayes algorithm. Naïve Bayes is a method that can be used to classify public opinion sentiment. The results of sentiment analysis research using naïve Bayes produce an accuracy rate of 0.82%. Then the model was tested to read public opinion on Twitter from 5 October 2022 to 27 October 2022. The results for Twitter user sentiment on the keyword Covid 19 were dominated by positive sentiment with a percentage of 85%. And sentiment with the keyword Omicron is still dominated by positive sentiment with a percentage of 49%. It was stated that the results of the classification on data for October 2022 meant that people were much more optimistic about the disappearance of the Omicron virus. Henceforth this research can be improved by adding data or using a different algorithm or implementing an existing algorithm

Keywords: Covid-19, Omicron, Social Media, Naïve bayes

1. PENDAHULUAN

Virus COVID-19 terus bermutasi membentuk varian baru. Varian terbaru yang terdeteksi, yaitu varian Omicron yang dikenal dengan B.1.1.529. Varian ini pertama kali dilaporkan dari Afrika Selatan pada 24 November 2021 dan telah menyebar ke seluruh dunia. Dan saat ini World Health Organization telah mengklasifikasikan varian Omicron sebagai VOC. Klasifikasi ini didasarkan pada temuan dari sejumlah besar mutasi dalam varian ini dan beberapa di antaranya menjadi perhatian. Penelitian pendahuluan menunjukkan peningkatan risiko infeksi ulang dengan varian ini dibandingkan dengan varian VOC lainnya. Apalagi varian ini juga menunjukkan pertumbuhan yang pesat. Berdasarkan hal tersebut varian ini akan memberikan dampak yang merugikan secara epidemiologi.

Bahkan kini sudah banyak negara yang mendeteksi varian yang masuk kategori 'variant of Concern' tersebut. Kategori VOC diartikan sebagai varian virus Corona yang diduga mampu menyebabkan peningkatan penularan serta kematian dan bahkan dapat mempengaruhi efektivitas vaksin[1].

Dari pertama kemunculan berita tentang COVID-19 pada Desember 2019, media sosial berperan penting sebagai tempat informasi berkembang biaknya *novel Corona virus* dari mulai kebenaran sampai kebohongan telah meledak di berbagai platform media sosial, bahkan persebaran isu pandemi di media sosial lebih cepat dibandingkan pandemi penyakit itu sendiri. Omicron menyebabkan gejala 'sangat ringan', Salah satu gejalanya termasuk kelelahan dan tenggorokan gatal. [2].

Media sosial seperti *twitter* dapat memberikan keuntungan bagi masyarakat, misalnya di dunia bisnis, masyarakat dapat mengetahui tentang layanan, produk, yang bisa dinilai baik atau buruk melalui opini yang ditulis masyarakat di platform media sosial seperti *twitter*[3].

Beberapa penelitian mengenai analisis sentimen terdahulu, salah satunya dalam penelitian analisis sentiment pandemic covid-19, penelitian tersebut bertujuan untuk melakukan collecting data twitter secara real time atau streaming data, dimulai pada tanggal 6 april 2020 dan data yang didapat 4000 data tweet dengan kata kunci Covid-19 dan coronavirus. Hasil dari nilai polaritas analisa sentiment dalam kategori netral paling tinggi yaitu 58,94% untuk sentiment Covid-19 dan 55,10% untuk variable coronavirus dibanding polaritas positif dan negatif[4].

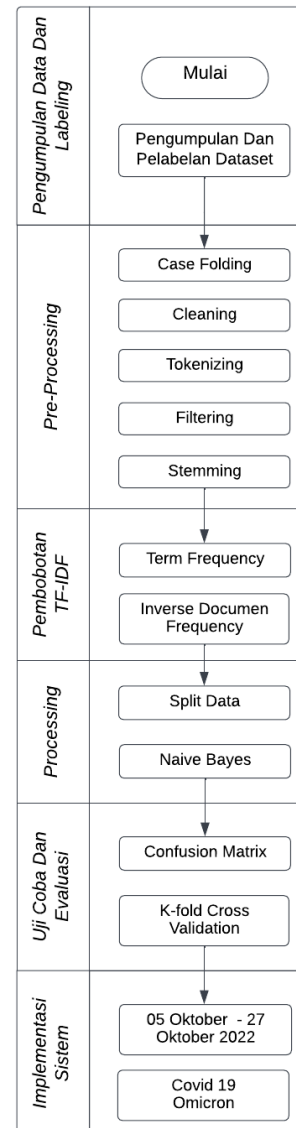
Pada penelitian lain menggunakan dua metode yaitu Naïve bayes dan KNN untuk mengklasifikasi opini publik mengenai Covid-19 pada twitter, penelitian ini berfokus untuk memberi gambaran pada masyarakat umum apakah Covid-19 cenderung ke opini positif atau negatif, serta membandingkan hasil akurasi dari dua metode KNN dan Naïve Bayes. Dengan hasil data yang didapat 1098 tweets, menggunakan kata kunci "COVID-19" pada tahun 2020. Hasil dari penelitian menunjukkan lebih banyak masyarakat yang beropini positif dengan 610 opini, dengan tingkat akurasi yang didapat Naïve bayes lebih tinggi sebesar 63.21% dibandingkan dengan metode KNN sebesar 58.10%[5].

Dalam penelitian ini memiliki tujuan utama yaitu membangun model yang dapat digunakan untuk megklasifikasi sentimen pada opini masyarakat di twitter dan youtube dengan metode Naïve Bayes.pembuatan model dengan menggunakan data 4233 yang diambil dari tanggal 22 februari 2022 sampai 9 juni 2022. Dan model akan digunakan untuk menganalisa sentimen pengguna twitter terhadap dua kata kunci yaitu Covid 19 dan Omicron, pada tanggal 5 Oktober 2022 sampai dengan 27 Oktober 2022.

2. BAHAN DAN METODE

Penelitian ini bertujuan untuk mengklasifikasikan sentiment masyarakat terhadap omicron pada kumpulan tweets ke dalam Positif, Negatif dan Netral. Proses klasifikasi dilakukan menggunakan algoritma Naïve Bayes. Model akan diimplementasikan untuk menganalisa sentimen pengguna *twitter* menggunakan data dengan dua kata kunci yaitu *covid-19* dan *omicron*.

Berikut beberapa tahapan dalam penelitian dapat dilihat pada gambar 1.



Gambar 1. Tahapan Penelitian

A. Pengumpulan data

Pengumpulan data dari penelitian ini yaitu data yang diunduh dari twitter berupa tweet beserta atribut-atributnya dan komentar dari youtube dengan menggunakan bantuan tools netlytic. Data dari twitter dan youtube yang diambil secara acak mulai dari tanggal 22 februari 2022 sampai 9 juni 2022 di media sosial dengan data yang digunakan sebanyak 4233 data, kemudian data yang akan dianalisa diambil dari twitter pada tanggal 5 Oktober 2022 sampai 27 Oktober 2022, menggunakan 2 kata kunci *Covid 19* dan *Omicron* yang berjumlah 5000.

B. Pelabelan Data

Pelabelan data yang dibuat sebagai model dilakukan secara manual terhadap 4233 data, dengan tiga pembagian kelas yaitu positif, negatif dan netral. Proses pemberian labeling dilakukan untuk menentukan tweet tersebut termasuk ke dalam kelas positif yang berisi informasi atau opini yang mengandung saran dan masukan mengenai virus Omicron. Sedangkan kelas negatif berisi keluhan, kalimat sindiran, kritik, dan cerminan emosi negatif seperti amarah, kesal, dan kecewa, dan netral berisi berita atau tweet yang tidak relevan dengan topik pembahasan. Setelah data diberi label secara manual maka data harus melalui proses validasi ahli pada penelitian ini data divalidasi oleh Dosen Ilmu Komunikasi di Universitas Darussalam Gontor yang Bernama Nur Aini Shofiya Asy'Ari, M.I.Kom. untuk contoh dari hasil pelabelan seperti pada tabel 1.

Tabel 1. Data label

No	Tweet	label
1	HAHAHAHA bangsuy! fix ini omicron bisa nular lewat vidcall!!!	Negatif
2	semoga cepat sembuh yaaa, sesama omicron alumni harus semangat ðŸ™ªðŸ™ªðŸ™ª»ðŸ™ªðŸ™ªðŸ™ª»ðŸ™ªðŸ™ªðŸ™ª»» btw ada gejala kah? kalau aku tenggorokan sih ðŸ™ª²	Positif
3	Ga sikat gigi sih jorok	Netral

Untuk hasil pelabelan data yang dilakukakn secara manual dengan pembagian kelas positif, negatif, dan netral seperti pada tabel 2

Tabel 2. Jumlah data labeling

No	Nama kelas	Jumlah Data
1	Positif	1540
2	Negatif	1547
3	Netral	1146

C. Pre-processing

Pada tahapan ini dilakukan proses reduksi teks yang bertujuan membuang kata-kata atau term-term yang tidak memiliki kontribusi dan menghilangkan *noise* pada data,

dengan tujuan membuat data lebih terstruktur. Adapun tahap-tahap *Pre-processing*.

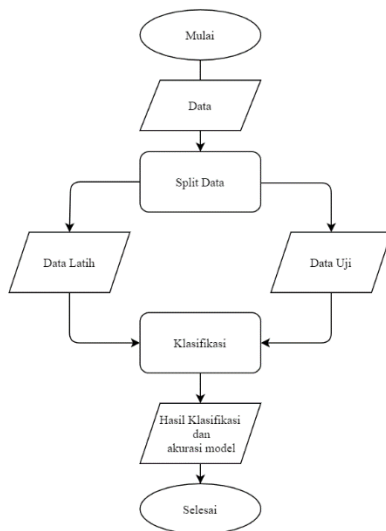
- Cleaning adalah tahap di mana karakter dan tanda baca yang tidak diperlukan dihapus dari teks. Cleaning bertujuan untuk mengurangi gangguan/noise pada dataset. Tahap-tahap pada cleansing ini adalah hapus url (<http://> atau <https://>), hapus username (@), hapus tanda, rt, hapus hashtag (#) dan tanda baca lainnya.
- Case folding adalah tahap merubah seluruh kata menjadi *huruf kecil* ataupun menjadi *huruf besar*[5].
- Tokenizing merupakan proses memecah kalimat menjadi kata-kata menjadi lebih berarti dan bermakna. Tokenisasi memisah setiap kata yang menyusun kalimat. Misalkan "kementerian pendidikan dan kebudayaan", setelah di tokenisasi menjadi "kementerian", "pendidikan", "dan", "kebudayaan"
- Filtering adalah tahap mengambil kata-kata penting dari hasil token dengan menggunakan algoritma stoplist (membuang kata kurang penting) atau wordlist (menyimpan kata penting) stopword adalah kata umum yang biasanya muncul dalam jumlah besar dan dianggap tidak memiliki makna. Contoh stopword dalam bahasa indonesia adalah "yang", "dan", "di", "dari", dll.
- Stemming merupakan tahapan untuk melakukan proses mengubah kata imbuhan infix(awal) maupun suffix(akhiran) menjadi sebuah kata dasar yang akan lebih mengandung sebuah makna untuk peroleh suatu informasi sehingga komentar akan menjadi lebih spesifik dalam pengkategorian contoh "menaikan", "naikan" akan ditransformasi menjadi kata "naik"[6]. Pada penelitian ini menggunakan library dari Satrawi yang dikembangkan oleh Nazief dan Andriani, perpustakaan Sastrawi juga menyediakan untuk berbagai Bahasa pemrograman contohnya PHP, Python, Java, C, GO dan Ruby[7].

D. Pembobotan kata

Pembobotan kata adalah sebuah suatu mekanisme untuk memberikan bobot atau nilai pada kata (term) terhadap frekuensi kemunculan suatu kata yang ada dalam dokumen. Terdapat beberapa metode untuk pembobotan kata (term weighting) diantaranya yaitu TF, TF-IDF, WIDF, dan TF-RF[8]. Term Frequency-Inverse Document Frequency merupakan metode pembobotan yang menggabungkan dua konsep yaitu Term Frequency dan Document Frequency. Term Frequency (TF) adalah konsep pembobotan dengan mencari seberapa sering (frekuensi) kemunculan suatu term dalam satu dokumen sedangkan IDF adalah semakin sedikit frekuensi kata muncul dalam dokumen, maka makin besar nilainya[9]. Dan dalam peneltian ini peneliti menggunakan TF-IDF.

E. Processing

Proses Processing merupakan proses pengelompokan data tertentu ke dalam beberapa kelas yang telah ditentukan, pada tahap penelitian ini akan menggunakan algoritma *Naïve Bayes classifier (NBC)*. *NBC* adalah salah satu metode untuk mengklasifikasi data dengan probabilitas sederhana yang mengaplikasikan teorema Bayes dengan karakter independen yang tinggi. Metode *Naïve Bayes* adalah metode klasifikasi text mining yang digunakan dalam analisis sentimen. Metode ini berpotensi bagus untuk klasifikasi presisi dan data komputasi. *Naïve Bayes* banyak digunakan dalam teknik klasifikasi terutama Twitter menggunakan beberapa metode seperti *Unigram Naïve Bayes*, *Multinomial Naïve Bayes*, dan *Maximum Entropy Classification*. Fitur utama klasifikasi *Naïve Bayes* untuk mendapatkan hipotesis yang kuat dari setiap kondisi atau peristiwa. Perhitungan kategori probabilitas pada *Naïve* menggunakan pendekatan algoritma Bayes dengan menggunakan persamaan. Metode ini sesuai untuk mengklasifikasi banyak data dengan cepat dan memiliki akurasi yang tinggi[10]. Pada gambar 2 dibawah ini menunjukkan proses klasifikasi menggunakan metode *naïve bayes*.



Gambar 2. Flowchart Naïve Bayes

Gambar diatas menjelaskan bahwa tahap pertama adalah memasukan data yang telah melewati pre-processing. Yang kemudian data dibagi menjadi dua bagian yaitu data testing dan data training. Data testing akan menjadi referensi bagi algoritma NB untuk mengklasifikasikan data. Data testing untuk menguji hasil pembelajaran dari algoritma *Naïve Bayes*. Hasil yang didapatkan berupa model yang dapat digunakan untuk mengklasifikasi dan nilai akurasinya.

$$P(X) = \frac{P(X|H)P(H)}{P(X)} \quad (1)$$

- X = data dengan kelas tidak dikenal
- H = hipotesis data X adalah kelas khusus
- P(H|X) = probabilitas hipotesis H didasarkan pada kondisi X
- P(H) = probabilitas hipotesis H
- P(X|H) = probabilitas hipotesis X didasarkan pada kondisi H
- P(X) = probabilitas X [6].

F. Evaluasi

Pada tahap ini selanjutnya melakukan evaluasi performa menggunakan *Confusion Matrix* untuk mengukur nilai performa berupa akurasi, presisi dan recall. Metode evaluasi ini berguna untuk menilai kinerja dari model klasifikasi yang sudah dibuat. Pada dasarnya metode ini membandingkan hasil label testing dengan label training. *Confusion matrix* yang digunakan dalam penelitian ini memiliki dimensi 3x3. Dimensi tersebut sesuai pembagian yang telah di tentukan yaitu positif, negatif dan netral. Dapat dilihat sebagai contoh pada gambar 3.

		Asli		
		Negatif	Netral	Positif
Asli	Negatif	True Negatif	False Negatif	False Negatif
	Netral	False Netral	True Netral	False Netral
	Positif	False Positif	False Positif	True Positif
		Negatif	Netral	Positif
		Prediksi		

Gambar 3. Model Confusion Matrix

Perhitungan nilai akurasi *Confusion matrix* berdimensi 3x3 berdasarkan[10] **Error! Reference source not found.** diperoleh dengan persamaan sebagai berikut:

$$\begin{aligned}
 & \text{Akurasi} \\
 &= \frac{PP + NgNg + NtNt}{PP + PNg + PNt + NgNg + NgP + NgNt + NtNt + NtP + NtNg}
 \end{aligned}$$

Keterangan:

- PP = Kategori Positif yang terprediksi Positif (Benar).
- PNg = Kategori Positif yang terprediksi Negatif (Salah).
- PNt = Kategori Positif yang terprediksi Netral (Salah).
- NgNg = Kategori Negatif yang terprediksi Negatif (Benar).
- NgP = Kategori Negatif yang terprediksi Positif (Salah).

- NgNt = Kategori Negatif yang terprediksi Netral (Salah)
- NtNt = Kategori Netral yang terprediksi Netral (Benar)
- NtP = Kategori Netral yang terprediksi Positif (Salah).
- NtNg = Kategori Netral yang terprediksi Negatif (Salah)[11].

G. Validasi

Pada tahap proses validasi model ini peneliti akan menggunakan metode *K-Fold Cross Validation* dalam melakukan validasi. *Cross Validation* merupakan metode yang sama dengan metode random subsampling dimana dataset yang diambil tidak tumpang tindih. Dalam k-fold cross validation set pelatihan dipartisi ke dalam subset k yang sama disetiap subsetnya. Data yang terdapat pada setiap partisi merupakan data acak yang terdapat pada data latih, kemudian model tersebut dilatih dengan bagian himpunan k dan diterapkan kepada subset lainnya[12]. Penelitian ini menggunakan 10 – k-fold untuk validasi. Contoh Ilustrasi pembagian data dapat dilihat pada Gambar 4.

No	Pembagian Dataset									
1	1	2	3	4	5	6	7	8	9	10
2	1	2	3	4	5	6	7	8	9	10
3	1	2	3	4	5	6	7	8	9	10
4	1	2	3	4	5	6	7	8	9	10
5	1	2	3	4	5	6	7	8	9	10
6	1	2	3	4	5	6	7	8	9	10
7	1	2	3	4	5	6	7	8	9	10
8	1	2	3	4	5	6	7	8	9	10
9	1	2	3	4	5	6	7	8	9	10
10	1	2	3	4	5	6	7	8	9	10
Keterangan										
	Nomor Pengujian									
	Data Test									
	Data Training									

Gambar 4. Ilustrasi *K-Fold Cross Validation*

3. HASIL DAN PEMBAHASAN

Penelitian ini bertujuan melakukan pengklasifikasian pada data tweet ke dalam kelas-kelas yang telah ditentukan. Pada penelitian ini kelas dibagi menjadi tiga kelas yaitu positif, negatif dan netral. Sebelum mengklasifikasikan data, data harus melewati proses preprocessing, *Preprocessing terdiri dari case folding, tokenizing, filtering dan stemming*. Seperti yang telah dijelaskan diatas. Tujuan dari Pre-processing adalah menjadikan data dapat tersusun dan mudah dimengerti oleh mesin pembelajaran. Setelah data melalui tahapan preprocessing data bisa melanjutkan ke tahapan

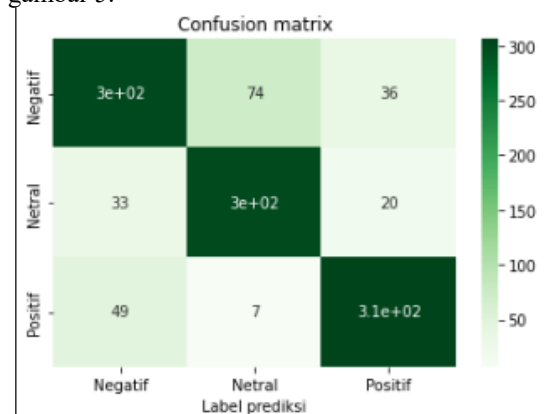
penerapan metode algoritma klasifikasi. Pembuatan model klasifikasi menggunakan fitur Google yaitu Google Colab. Dalam pembuatan model dibagi menjadi dua yaitu data testing dan data latih untuk data latih sebesar 0.8 dan data uji sebesar 0.2. berikut beberapa contoh hasil dari *Preprocessing* dapat dilihat pada tabel 3.

Tabel 3. Hasil preprocessing

Dok Tweet	label
ha bangsuy fix ini omicron	
1. bisa nular lewat vidcal	negatif
semoga cepat sembuh ya sesama omicron alumni harus semangat btw ada gejala kah kalau aku	
2. tengorokan sih	positif
3. ga sikat gi sih jorok	netral

1) Pembuatan model menggunakan metode naïve bayes

Metode klasifikasi naïve bayes digunakan untuk membuat model yang dapat menentukan apakah tweet masuk ke dalam salah satu dari tiga kelas tersebut. Model ini memanfaatkan nilai probabilitas dari setiap tweet untuk masuk kedalam kategori kelas yang telah ditentukan[13]. Proses klasifikasi dapat dilaksanakan apabila data yang digunakan telah melalui proses pembobotan dan pada penelitian ini menggunakan *TF/IDF*. Data yang telah melewati proses klasifikasi kemudian di evaluasi menggunakan *Confusion Matrix*. Dapat dilihat pada gambar 5.



Gambar 5. Hasil *Confusion Matrix*

Kemudian di validasi menggunakan *k-fold cross validation*. Pada penelitian ini menggunakan metode naïve bayes mendapatkan tingkat akurasi sebesar 0.82. Sedangkan untuk validasi menggunakan *K-fold cross validation* mendapatkan nilai rata-rata 0.79 dan hasil tertinggi 0.83.

Hasil dari validasi *K-fold cross validation* untuk model yang sudah dibuat menggunakan algoritma *naïve bayes* dengan 10 k-fold. Dengan rincian seperti pada tabel 4.

Tabel 4. Rincian K-fold

NO	K-fold
1	0.80487805
2	0.80487805
3	0.79600887
4	0.7804878
5	0.81152993
6	0.79600887
7	0.7827051
8	0.83333333
9	0.74666667
10	0.78
AVG	0.793649667405765

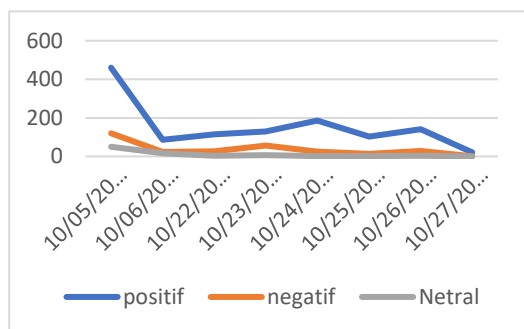
2) Implementasi model

Setelah melewati pembentukan model menggunakan metode naive bayes, maka langkah selanjutnya adalah implementasi data tweet dari tanggal 5 Oktober sampai dengan 27 Oktober 2022, menggunakan 2 kata kunci yang berjumlah 5000 data riil dengan rincian pada Tabel 5 menggunakan metode algoritma naive bayes.

Tabel 5. Jumlah data per kata kunci

Kata kunci	Jumlah data
Covid 19	2500
Omicron	2500

Presentase sentiment pengguna twitter pada kata kunci covid 19 dan omicron pada tanggal 5 Oktober sampai 27 Oktober 2022 dapat dilihat pada gambar 6.



Gambar 6. Tren sentimen per hari 5 Oktober sampai 27 Oktober 2022

3) Hasil analisis Sentimen

Hasil penelitian sentiment analisis menggunakan naive bayes menghasilkan tingkat akurasi sebesar 0.82%. Hasil dari pelatihan model kemudian model diuji untuk membaca opini public di twitter dan youtube dari tanggal 5 oktober 2022 sampai 27 oktober 2022, untuk hasil sentiment pengguna twitter pada kata kunci Covid 19 didominasi oleh sentiment positif dengan presentase 85%. Dan untuk sentiment dengan kata kunci Omicron masih didominasi oleh sentiment positif dengan presentase

49%. Dikarenakan adanya penurunan pada bulan oktober sehingga masyarakat jauh lebih optimis akan menghilangnya virus omicron. Dengan data pada tanggal 5 Oktober sampai 27 Oktober 2022.

4. KESIMPULAN

Berdasarkan dari hasil pengujian yang telah dilakukan, pengujian algoritma naive bayes terhadap opini masyarakat mengenai virus omicron pada sosial media twitter dan youtube dengan menggunakan pembagian tiga label, mendapat akurasi 0.82 dan di validasi *K-fold cross validation* mendapat rata-rata akurasi 0.79. Dengan hasil sentiment pengguna twitter pada kata kunci Covid 19 didominasi oleh sentiment positif dengan presentase 85%. Dan untuk sentiment dengan kata kunci Omicron masih didominasi oleh sentiment positif dengan presentase 49%. Disebutkan dari hasil klasifikasi pada data bulan oktober 2022 berarti masyarakat jauh lebih optimis akan menghilangnya virus omicron sehingga masyarakat jauh lebih tenang. Bagi peneliti semoga hasil dari penelitian ini dapat memberikan sumbangan pemikiran maupun sebagai masukan bagi peneliti lain terkait analisis sentiment untuk klasifikasi kedalam opini positif, negatif, dan netral. manfaat bagi pemerintah, hasil penelitian ini dapat memberikan informasi mengenai opini masyarakat terkait covid19 pada akhir tahun 2022. sehingga Pemerintah dapat mengambil langkah selanjutnya terkait peraturan terbaru kepada masyarakat mengenai covid19. Untuk selanjutnya penelitian ini dapat ditingkatkan dengan menambah data atau menggunakan algoritma yang berbeda ataupun implementasi pada algoritma yang sudah ada.

5. DAFTAR PUSTAKA

- [1] D. Satria, "Analisis dan Prediksi Kasus positif Covid-19 Varian Omicron dengan menggunakan Perbandingan Metode Backpropagation dan Metode Kalman Filter di Indonesia," *Adil J. Huk. STIH YPM*, vol. 4, no. 1, pp. 39–48, 2022, [Online]. Available: <https://adil.stihypm.ac.id/index.php/ojs/article/view/49>
- [2] A. Kumar, S. U. Khan, and A. Kalra, "COVID-19 pandemic: A sentiment analysis," *Eur. Heart J.*, vol. 41, no. 39, pp. 3782–3783, 2020, doi: 10.1093/eurheartj/ehaa597.
- [3] Normah, B. Rifai, S. Vambudi, and R. Maulana, "Adopsi Algorithm Support Vector Machine untuk Analisis Sentimen Larangan Mudik Lebaran 2020 pada Twitter," *J. Tek. Komput. AMIK BSI*, vol. 8, no. 2, pp. 174–180, 2022, doi: 10.31294/jtk.v6i2.8127.
- [4] A. K. Fauziyyah and D. H. Gautama, "Analisis Sentimen Pandemi Covid-19 Pada Streaming Twitter Dengan Text Mining Python," *J. Ilm. Sinus*, no. 2, pp. 31–42, 2020, doi: <http://dx.doi.org/10.30646/sinus.v18i2>.
- [5] M. Syarifuddin, "ANALISIS SENTIMEN OPINI PUBLIK MENGENAI COVID-19

- PADA TWITTER MENGGUNAKAN METODE NAÏVE BAYES DAN KNN,” *Inti Nusa Mandiri*, vol. 15, no. 1, pp. 23–28, 2020.
- [6] 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.
- [7] M. A. Rosid, A. S. Fitriani, I. R. I. Astutik, N. I. Mulloh, and H. A. Gozali, “Improving Text Preprocessing for Student Complaint Document Classification Using Sastrawi,” *IOP Conf. Ser. Mater. Sci. Eng.*, vol. 874, no. 1, 2020, doi: 10.1088/1757-899X/874/1/012017.
- [8] A. Deolika, K. Kusriani, and E. T. Luthfi, “Analisis Pembobotan Kata Pada Klasifikasi Text Mining,” *J. Teknol. Inf.*, vol. 3, no. 2, p. 179, 2019, doi: 10.36294/jurti.v3i2.1077.
- [9] A. Bayhaqy, S. Sfenrianto, K. Nainggolan, and E. R. Kaburuan, “Sentiment Analysis about E-Commerce from Tweets Using Decision Tree, K-Nearest Neighbor, and Naïve Bayes,” *2018 Int. Conf. Orange Technol. ICOT 2018*, no. November 2019, 2018, doi: 10.1109/ICOT.2018.8705796.
- [10] D. Normawati and S. A. Prayogi, “Implementasi Naïve Bayes Classifier Dan Confusion Matrix Pada Analisis Sentimen Berbasis Teks Pada Twitter,” *J-SAKTI (Jurnal Sains Komput. dan Inform.*, vol. 5, no. 2, pp. 697–711, 2021.
- [11] T. E. Hidayat, M. A. Rosid, and I. R. I. Astutik, “Analisa Sentimen Masyarakat Tentang Rencana Pemindahan Ibukota Negara Dengan Metode Naïve Bayes,” *JOINCS (Journal Informatics, Network, Comput. Sci.*, vol. 2, no. 1, pp. 1–12, 2020.
- [12] D. Berrar, “Cross-validation,” *Encycl. Bioinforma. Comput. Biol. ABC Bioinforma.*, vol. 1–3, no. January 2018, pp. 542–545, 2018, doi: 10.1016/B978-0-12-809633-8.20349-X.
- [13] U. Verawardina, F. Edi, and R. Watrianthos, “Analisis Sentimen Pembelajaran Daring Pada Twitter di Masa Pandemi COVID-19 Menggunakan Metode Naïve Bayes,” *Media Inform. Budirma*, vol. 5, pp. 157–163, 2021, doi: 10.30865/mib.v5i1.2604.