

ANALISIS SENTIMEN INSTAGRAM VAKSINASI MASA PANDEMI COVID-19 MENGUNAKAN METODE NAÏVE BAYES

Dwi Winarti, Faculty of Science and Technology, Universitas PGRI Yogyakarta, Fanny Aura Salsabila, Faculty of Science and Technology, Universitas PGRI Yogyakarta, Fadia Ayu Cahyani, Faculty of Science and Technology, Universitas PGRI Yogyakarta, Nurirwan Saputra, Faculty of Science and Technology, Universitas PGRI Yogyakarta, Meilany Nonsi Tentua, Faculty of Science and Technology, Universitas PGRI Yogyakarta, Ahmad Riyadi, Faculty of Science and Technology, Universitas PGRI Yogyakarta

Abstrak— Pandemi Covid-19 merupakan virus yang sudah menyebar ke seluruh dunia. Pemerintah juga telah meminimalisir penyebaran virus dengan melakukan Pembatasan Sosial Berskala Besar (PSBB), penerapan lockdown, larangan mudik saat hari raya idul fitri dan lain sebagainya. Pemerintah juga telah memanfaatkan media sosial sebagai informasi dan pelayanan untuk masyarakat. Pemerintah juga tidak hanya diam ketika banyak dampak negative terhadap masyarakat maka pemerintah mengambil tindakan vaksinasi. Tanggapan masyarakat terhadap vaksinasi cukup menarik dan beragam di Instagram baik positif maupun negatif. Maka penelitian ini bertujuan untuk menganalisis sentimen terhadap vaksinasi pandemic Covid-19 menggunakan metode naïve bayes. Berdasarkan hasil pengujian analisis sentimen pengguna Instagram vaksinasi dengan tokenisasi Quadgram mendapatkan nilai akurasi sebesar 75,9124 %.

Kata Kunci : Covid-19, Naïve Bayes, Vaksin, Instagram.

Abstract — *The Covid-19 pandemic that has spread throughout the world. The government has also minimized the spread of the virus by implementing Large-Scale Social Restrictions (PSBB), implementing a lockdown, banning going home during Eid al-Fitr and so on. The government has also utilized social media as information and services for the community. The government also does not remain silent when there are many negative impacts on society, so the government takes vaccination action. The public's response to the vaccination program is quite interesting and varied on Instagram, both positive and negative. So this study aims to analyze sentiment towards the Covid-19 pandemic vaccination program using the Naïve Bayes method. Based on the results of sentiment analysis testing, the Instagram user with Quadgram tokenization obtained an accuracy value of 75.9124%.*

Keyword : COVID-19, Naïve Bayes, Vaccines, Instagram.

I. PENDAHULUAN

Pandemi Covid-19 merupakan corona virus yang menyebar ke Indonesia membawa banyak perubahan

baik dari sosial masyarakat maupun dari kehidupan ekonomi. Pemerintah telah mencoba mengatasi wabah dengan mempertimbangkan serangkaian tindakan, tidak semuanya sesuai dengan opini publik. Selama ini pesatnya pertumbuhan jumlah kasus secara global telah menimbulkan kepanikan, ketakutan, dan kecemasan di kalangan masyarakat.

Upaya pemerintah Indonesia di antaranya dengan memberlakukan kebijakan protokol kesehatan, Pembatasan Sosial Berskala Besar (PSBB) hingga vaksinasi masyarakat untuk menekan penyebaran Covid-19. Vaksin COVID-19 kini telah tiba di Indonesia. Pengujian keamanan dan efektivitasnya telah dan sedang dilakukan pihak terkait. Penegasan Presiden mengenai vaksinasi kepada seluruh masyarakat dengan gratis. Bahkan, untuk membangun kepercayaan publik, Presiden Jokowi memberikan contoh lebih dulu untuk divaksin.

Analisis sentimen berada di antara berbagai bidang penelitian lainnya seperti Data Mining, Natural Language Processing (NLP) dan *Machine learning* untuk melakukan ekstraksi sentimen terhadap sebuah isi kalimat. Analisis sentimen dilakukan untuk mengelompokkan dokumen, kalimat, atau pendapat berdasarkan polaritas teks di dalamnya. Polaritas tersebut merupakan pendapat yang memiliki aspek positif, negatif atau netral.

Klasifikasi dengan analisis sentimen membantu memberikan masukan dan tanggapan dari pelanggan dengan cepat. Naive Bayes merupakan teknik klasifikasi

yang paling sering digunakan untuk klasifikasi data yang disebut dengan Naive Bayes Classifier (NBC). Pemilihan algoritma Naive Bayes ini karena sangat cocok untuk short data text dan sederhana namun memiliki nilai akurasi yang tinggi dalam pengklasifikasian data teks. Metode *Naive Bayes* banyak digunakan di berbagai penelitian tentang analisis dokumen tekstual dan merupakan algoritma populer karena kemudahan penggunaannya serta telah terbukti memuaskan di banyak domain penelitian[1].

II. LANDASAN TEORI

a. Analisis Sentimen

Analisis sentimen atau opinion mining mencakup berbagai macam penyebutan yang berbeda-beda diantaranya analisis sentimen, opinion mining, opinion extraction, sentiment mining, subjectivity analysis, affect analysis, emotion analysis, dan review mining [2]. Analisis sentimen melakukan pembelajaran kepada komputer sehingga komputer mampu dalam menentukan sebuah kalimat yang dituliskan oleh seseorang merupakan kalimat positif, netral atau negatif. Analisis Sentimen sudah banyak dimanfaatkan dalam berbagai penelitian, diantaranya terkait manajemen reputasi tokoh masyarakat [3], [4], respon masyarakat dalam teknologi baru [5], Respon mahasiswa terkait kualitas pembelajaran [6], Termasuk Analisis Sentimen masyarakat terkait review sebuah film [7].

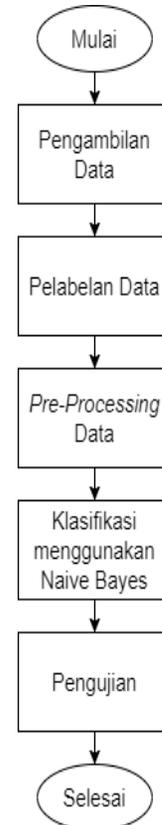
b. Naïve Bayes

Naïve Bayes merupakan salah satu metode klasifikasi yang menerapkan probabilitas dan statistika menggunakan pendekatan Bayes, penggunaan teorema Bayes pada algoritma *Naïve Bayes* adalah dengan menggabungkan *prior probability* dan *conditional probability* dalam suatu rumus yang dapat digunakan untuk menghitung probabilitas dari setiap kemungkinan klasifikasi[8]. *Naïve Bayes* cukup populer digunakan untuk mengklasifikasikan sentimen seseorang, diantaranya review BMKG [9], review toko online [10], respon terkait Covid-19 [11], dan masih banyak lagi lainnya. *Naive Bayes Classifier* merupakan salah satu metode *machine learning* yang menggunakan perhitungan probabilitas. Keuntungan dari penggunaan metode *Naïve Bayes* adalah hanya membutuhkan sejumlah kecil data pelatihan untuk memperkirakan parameter (sarana dan varians dari variabel-variabel) yang diperlukan untuk klasifikasi. Karena variabel independen diasumsikan, hanya varian dari variabel untuk masing-masing kelas harus ditentukan dan tidak seluruh matriks kovariansi [12]. Data latih yang sudah

melewati *pre-processing* akan menjadi bahan pembelajaran untuk proses klasifikasi data uji menggunakan algoritme *Naïve Bayes*[13].

III. METODE

Metode yang dilakukan pada penelitian ini dengan mengambil dataset komentar mengenai “Komentar vaksinasi yang telah dilaksanakan di Indonesia”. Penelitian ini menganalisa sentimen dengan menggunakan metode Naive Bayes untuk menentukan akurasinya.



Gambar 1. *flowchart* metodologi

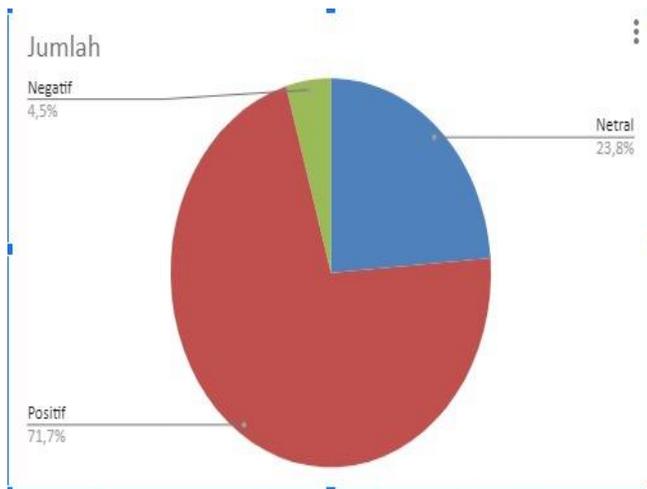
1. Pengumpulan Data

Proses ini dilakukan dengan cara mengumpulkan komentar dari postingan akun Instagram milik Presiden RI yaitu Jokowi, yang mulanya dicari melalui “#vaksinasi”. Setelahnya melakukan web *scraping* untuk mengambil komentar pada postingan tersebut, dan diperoleh sebanyak 685 komentar untuk sampel penelitian. . Pengumpulan data komentar Instagram dari data tersebut kemudian dilakukan pelabelan secara manual dengan 3 pembagian pelabelan yakni positif, negatif dan netral. Berikut adalah hasil dari label sentimen komentar vaksin COVID-19 ditunjukkan pada Gambar berikut.

ariboy1988 bukan hanya mall aja sarng sebar virus corona	Netral
auracamilla semangat terus indonesia	Positif
affandi3645 moga pandemi ini segera lalu dan tidak ada lagi yang papar covid lagi di negara kita indonesia aamin	Positif
adi soe alfonsus raharja trus gimana nasib karyawan pak blom juga baik jam kerja nya di jawa timur	Netral
lestaris3ndra semua gantung anda	Netral
ini ratno terimakasih atas semua tinda para tinda baik itu yang organisir dari kelompok tentu maupun orang yang tela	Positif
55555 fmp pakde please segera tiada pcr utk jalan udara bagi yg udah vaksin 2 x utama daerah sulawesi tengah	Netral
mutiara99994 itu penting	Positif
yakobuswisnu moga indonesia raya lekas sembuh lahir kembali seperti kertas putih yang belum kena tinta	Positif
oziyusuf4merdeka	Netral
agathandmade boutique mampir dan bantu larisin yuk kakak2 sayang	Netral

Gambar 2. Pelabelan Manual

Berdasarkan hasil pengujian terdapat 685 data latih Sebagian besar di Instagram memiliki sikap positif (71,7 %), negatif (4,5 %) dan netral (23,8 %). Jika ditotalkan maka ada sebanyak yakni 491 komentar sentiment dengan positif, 31 komentar dengan sentimen negatif dan 163 komentar dengan sentimen netral.



Gambar 3. Grafik Persentase Pelabelan

2. Pelabelan Manual

Proses ini dilakukan secara manual untuk mengidentifikasi mana komentar yang mengandung negatif, positif dan netral. Pada proses ini peneliti menggunakan pedoman label untuk menentukan label setiap komentar pada data, sebagai berikut:

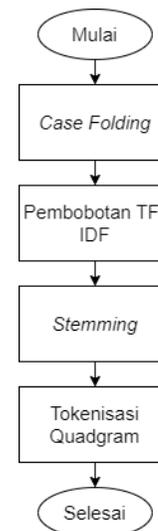
- a. Komentar Negatif merupakan komentar dari masyarakat yang menolak divaksin dan tidak setuju dengan kebijakan pemerintah mewajibkan Vaksinasi COVID-19 kepada warganya, maupun tidak setuju dengan apa yang dilakukan pemerintah dalam mengatasi pandemi ini.
- b. Komentar Positif merupakan komentar dari masyarakat menerima vaksin dan setuju dengan kebijakan pemerintah terkait mewajibkan

Vaksinasi COVID-19 kepada warganya, maupun setuju dengan apa yang dilakukan pemerintah dalam mengatasi pandemi ini.

- c. Komentar Netral merupakan komentar dari masyarakat yang bersifat Netral, tidak berisi komentar yang setuju maupun tidak terhadap kebijakan pemerintah dalam mewajibkan Vaksinasi COVID-19 kepada warganya.

3. Preprocessing Data

Preprocessing data langkah awal untuk mengolah data teks yang tidak teratur menjadi bentuk data yang sesuai dengan kebutuhan penelitian.



Gambar 4. Preprocessing Data

Maka metode yang digunakan pada pengolahan data sebagai berikut:[14]

a. Case Folding

Case Folding adalah mengubah semua huruf menjadi huruf kecil, karena komputer akan menganggap sebuah kata menjadi berbeda apabila salah satu hurufnya berbeda besar dan kecilnya, karena komputer mengenali setiap karakter huruf ke dalam biner dan huruf kecil maupun besar memiliki angka biner yang berbeda. Dengan adanya perbedaan ini, sehingga kata yang seharusnya sama menjadi berbeda dan menimbulkan variasi kata yang lebih banyak, sehingga dengan adanya Case Folding akan meminimalisir variasi kata yang akan dipelajari komputer.

b. Pembobotan TF-IDF

Metode TF-IDF merupakan cara untuk memberikan bobot ada hubungan sebuah kata

(term) dengan sebuah dokumen. Metode ini menggabungkan dua konsep untuk menghitung bobot, yaitu frekuensi kemunculan suatu kata dalam dokumen tertentu dan frekuensi kebalikan dari dokumen yang mengandung kata tersebut [15].

Gambar 5. Perhitungan TF-IDF

c. *Stemming*

Stemming adalah tahapan mengubah kata menjadi kata dasar menurut kaidah bahasa Indonesia yang benar [8]. Proses ini dilakukan setelah proses pelabelan selesai. Proses akan dilakukan menggunakan jupyter notebook dengan python. Dengan menggunakan perintah untuk mengubah seluruh isi dataset komentar menjadi ke kata dasar dan stemming yang digunakan pada penelitian ini adalah Sastrawi Master [16]. Berikut contoh implementasi *Stemming*:

Tabel 1. Perbandingan Kata Saat *Stemming*

Sebelum	Setelah
Memperbaiki	baik
ditindas	tindas
terpapar	papar
perjuangan	juang

d. Tokenisasi

Tahap memotong *string input* berdasarkan kata yang menyusunnya serta membedakan karakter-karakter tertentu yang diperlukan sebagai pemisah kata atau bukan [1]. Proses pemisahan kalimat menjadi sebuah kata. Pada proses ini peneliti memilih quadgram atau pemisahan kalimat menjadi 4 suku kata untuk tokenisasi.

Tabel 2. Tokenisasi Quadgram

Kalimat	Trigram
Bukan hanya mall aja sarng penyebaran virus Corona	“Bukan hanya mall aja”, ”hanya mall aja sarng”, ”mall aja sarng”, ”aja sarng”, ”aja sarng penyebaran virus”, ”sarng penyebaran virus corona”

IV. HASIL DAN PEMBAHASAN

a. Hasil

Diketahui pada pengklasifikasian dataset ini menghasilkan *accuracy*, *precision* dan *recall* sebagai berikut

Tabel 3. Prediksi Akurasi Quadgram

<i>Accuracy</i>	<i>Precision</i>	<i>Recall</i>
0.7591	0.758	0.759
75.91%	75.80%	76%

Pada penelitian ini prediksi pada tokenisasi quadgram menghasilkan akurasi sebanyak 75.91%, *precision* 75.80% dan *recall* 76%.

b. Pembahasan

Dengan menggunakan tokenisasi Quadgram, maka untuk mendapatkan makna suatu kalimat terbatas harus kombinasi 4 kata, sedangkan makna dari suatu kalimat sudah bisa diambil dari satu sampai 2 kata, sehingga akurasi yang dihasilkan tidak cukup tinggi.

V. KESIMPULAN

5.1 Kesimpulan

Maka dengan penelitian ini mendapatkan dataset berisikan 685 data komentar dari Instagram mengenai “Vaksinasi yang telah dilaksanakan di Indonesia” dengan menggunakan *scraping*. Pada dataset tersebut diketahui mendapat 3 sentimen untuk dijadikan pedoman label yaitu: positif, negatif dan netral. Penggunaan metode klasifikasi Naive Bayes dapat menghasilkan *accuracy* sebanyak 75.91%, *precision* 75.80% dan *recall* 76%.

5.1 Saran

Saran pengembangan dari penelitian yang sudah dilakukan diantaranya adalah perlu adanya implementasi secara utuh analisis sentimen yang dikelola oleh pihak independent untuk kepentingan banyak hal yang mendapatkan akses khusus untuk dapat mengambil data sesuai dengan kata kunci tertentu dari berbagai macam sumber data.

DAFTAR PUSTAKA

- [1] R. D. Septiana, A. B. Susanto, And T. Tukiyyat, "Analisis Sentimen Vaksinasi Covid-19 Pada Twitter Menggunakan Naive Bayes Classifier Dengan Feature Selection Chi-Squared Statistic Dan Particle Swarm Optimization," *Jurnal Siskom-Kb (Sistem Komputer Dan Kecerdasan Buatan)*, Vol. 5, No. 1, Pp. 49–56, 2021, Doi: 10.47970/Siskom-Kb.V5i1.228.
- [2] N. Saputra, "Analisis Sentimen Dengan Preprocessing Kata (Sentiment Analisis With Lexicon Preprocessing)," *Jurnal Dinamika Informatika*, Vol. 7, No. 1, Pp. 45–57, 2018, Accessed: Oct. 26, 2020. [Online]. Available: <https://Jdi.Upy.Ac.Id/Index.Php/Jdi/Article/View/14>
- [3] N. Saputra, K. Nurbagja, And T. Turiyan, "Sentiment Analysis Of Presidential Candidates Anies Baswedan And Ganjar Pranowo Using Naive Bayes Method," *Jurnal Sisfotek Global*, Vol. 12, No. 2, Pp. 114–119, Sep. 2022, Doi: 10.38101/Sisfotek.V12i2.552.
- [4] N. Saputra, T. B. Adji, And A. E. Permanasari, "Analisis Sentimen Data Presiden Jokowi Dengan Preprocessing Normalisasi Dan Stemming Menggunakan Metode Naive Bayes Dan Svm Oleh," *Jurnal Dinamika Informatika*, Vol. 5, No. 1, 2015.
- [5] N. Saputra, "Analisis Sentimen Self-Driving Car Dengan Sentiment Confident Terbaik," In *Seminar Nasional Dinamika Informatika*, 2019, Pp. 40–44. [Online]. Available: <http://Prosiding.Senadi.Upy.Ac.Id/Index.Php/Senadi/Article/View/101>
- [6] N. Saputra, "Analisis Sentimen Mahasiswa Terhadap Universitas," In *Seminar Nasional Dinamika Informatika*, 2016. Accessed: Oct. 26, 2020. [Online]. Available: https://Nanopdf.Com/Download/Seminar-Nasional-Universitas-Pgri-Yogyakarta-2016-Isbn-978-3_Pdf
- [7] S. B. Putri, Y. N. Anisa, And N. Saputra, "Analisis Sentimen Film Kuliah Kerja Nyata (Kkn) Di Desa Penari Menggunakan Metode Naive Bayes," *Jusitik : Jurnal Sistem Dan Teknologi Informasi Komunikasi*, Vol. 5, No. 2, Pp. 22–26, Oct. 2022, Doi: 10.32524/Jusitik.V5i2.704.
- [8] W. Yulita *Et Al.*, "Analisis Sentimen Terhadap Opini Masyarakat Tentang Vaksin Covid-19 Menggunakan Algoritma Naive Bayes Classifier," *Jdmsi*, Vol. 2, No. 2, Pp. 1–9, 2021.
- [9] D. Darwis, N. Siskawati, And Z. Abidin, "Penerapan Algoritma Naive Bayes Untuk Analisis Sentimen Review Data Twitter Bmkg Nasional," *Jurnal Tekno Kompak*, Vol. 15, No. 1, Pp. 131–145, Feb. 2021, Doi: 10.33365/Jtk.V15i1.744.
- [10] F. V. Sari And A. Wibowo, "Analisis Sentimen Pelanggan Toko Online Jd.Id Menggunakan Metode Naive Bayes Classifier Berbasis Konversi Ikon Emosi," *Simetris: Jurnal Teknik Mesin, Elektro Dan Ilmu Komputer*, Vol. 10, No. 2, Pp. 681–686, Nov. 2019, Doi: 10.24176/Simet.V10i2.3487.
- [11] M. Syarifuddin, "Analisis Sentimen Opini Publik Mengenai Covid-19 Pada Twitter Menggunakan Metode Naive Bayes Dan Knn," *Inti Nusa Mandiri*, Vol. 15, No. 1, Pp. 23–28, Aug. 2020, Doi: 10.33480/Inti.V15i1.1347.
- [12] L. Oktasari *Et Al.*, "Text Mining Dalam Analisis Sentimen Asuransi Menggunakan Metode Naive Bayes Classifier," *Prosiding Seminar Sains Nasional Dan Teknologi*, Vol. 1, No. 1, Sep. 2016, Doi: 10.36499/Psnst.V1i1.1506.
- [13] B. Laurensz And Eko Sedyono, "Analisis Sentimen Masyarakat Terhadap Tindakan Vaksinasi Dalam Upaya Mengatasi Pandemi Covid-19," *Jurnal Nasional Teknik Elektro Dan Teknologi Informasi*, Vol. 10, No. 2, Pp. 118–123, 2021, Doi: 10.22146/Jnteti.V10i2.1421.
- [14] F. Septianingrum, J. H. Jaman, And U. Enri, "Analisis Sentimen Pada Isu Vaksin Covid-19 Di Indonesia Dengan Metode Naive Bayes Classifier," *Jurnal Media Informatika Budidarma*, Vol. 5, No. 4, P. 1431, 2021, Doi: 10.30865/Mib.V5i4.3260.
- [15] Merinda Lestandy, Abdurrahim Abdurrahim, And Lailis Syafa'ah, "Analisis Sentimen Tweet Vaksin Covid-19 Menggunakan Recurrent Neural Network Dan Naive Bayes," *Jurnal Resti (Rekayasa Sistem Dan Teknologi Informasi)*, Vol. 5, No. 4, Pp. 802–808, 2021, Doi: 10.29207/Resti.V5i4.3308.
- [16] "Github - Sastrawi/Sastrawi: [Inactive] High Quality Stemmer Library For Indonesian Language (Bahasa)." <https://Github.Com/Sastrawi/Sastrawi> (Accessed Jan. 31, 2023).



Sistem Cerdas.

Dwi Winarti, Lulus dari SMK Muhammadiyah 2 Bantul jurusan Rekayasa Perangkat Lunak (RPL) dan selanjutnya menyelesaikan pendidikannya pada tahun 2019. Dwi Winarti saat ini melanjutkan pendidikannya di Universitas PGRI Yogyakarta mengambil program studi S1 Informatika dan mengambil minat di bidang



Fanny Aura Salsabila, Lulus dari SMK Kesehatan Bantul jurusan Keperawatan dan selanjutnya menyelesaikan pendidikannya pada tahun 2019. Fanny Aura Salsabila saat ini melanjutkan pendidikannya di Universitas PGRI Yogyakarta mengambil program studi (Prodi) S1 Informatika dan mengambil minat di bidang Sistem Cerdas.



Fadia Ayu Cahyani, Lulus dari SMAN 1 Sedayu jurusan Bahasa dan Budaya menyelesaikan pendidikannya pada tahun 2019. Fadia Ayu Cahyani saat ini melanjutkan pendidikannya di Universitas PGRI Yogyakarta mengambil program studi S1 Informatika dan mengambil minat di bidang Sistem Cerdas,



Nurirwan Saputra, lulus S1 Teknik Informatika Universitas Islam Indonesia (UII) pada tahun 2011, Lulus S2 Magister Teknologi Informasi, Universitas Gadjah Mada (UGM) pada tahun 2015. Saat ini bekerja sebagai dosen Prodi

Informatika Universitas PGRI Yogyakarta (UPY). Bidang yang dijalani berkaitan dengan Sentiment Analysis