



Analisis Sentimen Pengguna Twitter Terhadap Vaksin Booster Menggunakan Metode Naïve Bayes Dan K-Nearst Neighbor

Alifia Septi Rizqiah¹, Abdul Rahman Kadafi²

^{1,2} Jl. Margonda Raya No.545, Indonesia

¹alifiaseptirizqiah@gmail.com, ²abdul.alk@nusamandiri.sch.id

Abstract- The outbreak of the Covid-19 pandemic disease, which stands for the corona virus disease-19, has hit all over the world, one of which is Indonesia with a large number of cases. The impact with many people infected with Covid-19 has made the government take steps to reduce the spread of Covid-19 in Indonesia, one of which is vaccination. Following the administration of dose one and dose two vaccinations that have been widely carried out by the public, the Ministry of Health issued a circular regarding the COVID-19 Advanced Dose (Booster) Vaccination. The implementation of booster vaccines raises various kinds of pro and contra opinions in the community, which can be found on social media, one of which is Twitter. Based on these problems, the researchers used the Naïve Bayes and K-Nearst Neighbor methods to analyze the sentiments of Twitter users towards the booster vaccine. Data grouping is 1000 data obtained from twitter with Naïve Bayes and K-Nearst Neighbor methods. The accuracy results obtained are 90.19% with the Naïve Bayes method and the accuracy results are 82.89% with the K-Nearst Neighbor (K-NN) method. The precision value obtained by the Naïve Bayes method is 95.23% for positive pred, which is higher than for negative pred, which is 86.16%.

Keywords- Covid-19, Vaccine, Sentiment Analysis, K-NN, Naïve Bayes

Abstrak- Wabah penyakit pandemi Covid-19 kependekan dari corona virus disease-19 yang melanda di seluruh dunia salah satunya Indonesia dengan jumlah kasus yang tidak sedikit. Dampak dengan banyak masyarakat yang terinfeksi Covid-19 membuat pemerintah mengambil langkah mengurangi penyebaran Covid-19 di Indonesia salah satunya pemberian vaksinasi. Setelah pemberian vaksinasi dosis satu dan dosis dua yang telah banyak dilakukan masyarakat, Kementerian Kesehatan menerbitkan surat edaran tentang Vaksinasi COVID-19 Dosis Lanjutan (Booster). Pelaksanaan vaksin booster menimbulkan berbagai macam pendapat pro dan kontra di masyarakat yang banyak ditemukan di media sosial salah satunya twitter. Berdasarkan permasalahan tersebut maka peneliti menggunakan metode Naïve Bayes dan K-Nearst Neighbor untuk analisis sentimen pengguna twitter terhadap vaksin booster. Pengelompokan data sebanyak 1000 data yang didapat dari twitter dengan metode Naïve Bayes dan K-Nearst Neighbor. Hasil akurasi yang didapatkan sebesar 90.19% dengan metode Naïve Bayes dan hasil akurasi sebesar 82.89% dengan metode K-Nearst Neighbor (K-NN). Nilai precision yang didapatkan dengan metode Naïve Bayes yaitu 95.23% untuk pred.positif lebih tinggi dibandingkan pred.negatif yaitu 86.16%.

Kata kunci- Covid-19, Vaksin, Analisis Sentimen, K-NN, Naïve Bayes

I. PENDAHULUAN

Wabah penyakit pandemi Covid-19 kependekan dari corona virus disease-19 yang melanda di seluruh dunia salah satunya Indonesia dengan jumlah kasus yang tidak sedikit. Menurut data dari laman peta sebaran covid.go.id [1] hingga 18 April 2022 jumlah kasus positif Covid-19 di Indonesia berjumlah total 6,04 juta kasus positif. Banyak kasus terkait dengan pandemi ini yang memakan korban jiwa yang tidak sedikit dengan terinfeksi virus Covid-19 sehingga harus membuat upaya untuk mencegah dampak tersebut semakin menyebar luas.

Vaksinasi merupakan pemberian vaksin (antigen) yang digunakan untuk merangsang sistem imun dalam tubuh. Berdasarkan data dari Kementerian Kesehatan [2] jumlah nasional masyarakat yang sudah vaksinasi per 19 April 2022 dengan vaksinasi dosis satu sudah mencapai 95,26% dan vaksinasi dosis dua sudah mencapai 78,30%. Vaksinasi dilakukan dengan pemberian dosis satu dan dosis dua pada masyarakat telah banyak dilakukan di banyak tempat sebagai upaya mempercepat menekan kasus

penyebaran Covid-19 dan terbentuknya kekebalan terhadap penyakit atau disebut herd immunity pada masyarakat [3].

Pemberian vaksinasi booster ini berdasarkan hasil studi yang menunjukkan terjadi penurunan antibodi dalam 6 bulan setelah mendapatkan vaksinasi Covid-19 dosis primer lengkap sehingga dibutuhkan pemberian dosis lanjutan atau booster untuk meningkatkan proteksi individu terutama pada masyarakat rentan. Pelaksanaan vaksin booster menimbulkan berbagai macam pendapat pro dan kontra di masyarakat yang banyak ditemukan di media sosial salah satunya twitter. Media sosial twitter banyak digunakan untuk memberikan opini salah satunya tentang wabah virus corona [4]. Upaya mengadakan suatu program agar efektif dan efisien dapat menggunakan media sosial twitter. Kesuksesan program vaksin dapat mendorong opini masyarakat untuk vaksin [3].

Dalam berbagai media sosial salah satunya twitter kita dapat mengetahui banyak opini seperti opini positif ataupun negatif dengan sebuah teknik yang digunakan



seperti text mining [5]. Text mining adalah teknik yang digunakan untuk memenuhi kebutuhan dalam proses klasifikasi dokumen dengan konten apapun. Salah satu teknik mengubah sebuah informasi yang merupakan sikap seseorang atau suatu kejadian dengan membagi polaritas dari sebuah teks adalah analisa sentimen. Pembagian tersebut untuk mengetahui teks tersebut bersifat positif, negatif atau netral [6]. Berdasarkan banyak opini yang berada pada twitter dapat diolah menjadi data untuk dijadikan analisis sentimen.

Metode klasifikasi data Naive Bayes memiliki teori kinerja yang sejenis dengan Decision Tree dan Neural Network. Naive bayes yang dikemukakan oleh ilmuwan Inggris Thomas Bayes untuk merupakan salah satu metode klasifikasi data yang memakai perhitungan probabilitas dan statistic [7]. Metode klasifikasi data K-Nearest Neighbor merupakan metode pengklasifikasian suatu objek penelitian bersumber dari gabungan data latih yang secara perhitungan memiliki jarak terkecil dengan objek data uji yang berbeda

Dalam penelitian terdahulu mengenai analisis sentimen tentang vaksinasi covid-19 menggunakan metode Support Vector Machine [8][9], metode Naive Bayes [4][5], metode K-Nearest Neighbor [8][10]. Menurut [11], metode klasifikasi data menggunakan Naive Bayes memiliki akurasi yang tinggi sehingga hasil dari prediksinya akurat untuk dijadikan metode dalam penelitian selanjutnya. Menurut penelitian [12] nilai akurasi dengan menggunakan metode K-Nearest Neighbor mencapai 87.00%.

Terkait metode yang digunakan untuk analisis sentimen oleh Tanthy Tawaqalia Widowati & Mujiono Sadikin dengan judul Analisis Sentimen Twitter Terhadap Tokoh Publik Dengan Algoritma Naive Bayes dan Support Vector Machine menghasilkan kesimpulan nilai akurasi dari metode naive bayes adalah 91.58% dan hasil akurasi dari SVM adalah 85.47% sehingga metode naive bayes lebih baik dalam memprediksi lebih akurat dan tepat [13]. Adapun metode penelitian untuk analisis sentimen oleh Dedi Darwis, Nery Siskawati dan Zaenal Abidin dengan judul Penerapan Algoritma Naive Bayes Untuk Analisis Sentimen Review Data Twitter BMKG Nasional menghasilkan kesimpulan analisis sentimen yang dilakukan dengan membuat kelas positif, netral dan negatif menggunakan algoritma naive bayes dengan tingkat akurasi yaitu 68.97% [7].

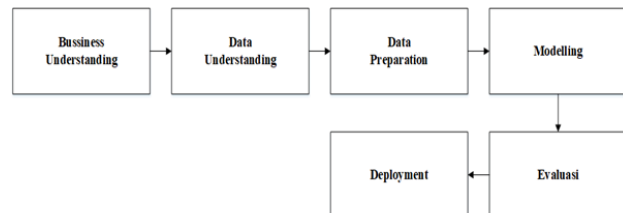
Penelitian yang dilakukan oleh M.Suyudi Alrajak, Iin Ernawati dan Ika Nurlaili dengan judul Analisis Sentimen Terhadap Pelayanan PT PLN Di Jakarta Pada Twitter Dengan Algoritma K-Nearest Neighbor (K-NN) menghasilkan kesimpulan bahwa algoritma K-NN dapat digunakan untuk membuat analisis sentimen khususnya pelayanan dalam perusahaan dengan hasil total nilai akurasi sebesar 89.4% [14]. Penelitian yang dilakukan oleh Blidex dan Jati Sasongko Wibowo dengan judul Analisis Sentimen Klasifikasi Tweet Vaksin Covid-19 Dengan Naive Bayes menghasilkan kesimpulan nilai akurasi yang dihasilkan sebanyak 96% sehingga metode naive bayes dapat digunakan untuk melakukan klasifikasi [15].

Penelitian yang dilakukan oleh Slamet Harry Ramdhani dan Muhammad Iwan Wahyudi dengan judul Analisis Sentimen Terhadap Vaksinasi Astra Zeneca Pada Twitter Menggunakan Metode Naive Bayes dan K-NN menghasilkan kesimpulan nilai akurasi yang diperoleh dengan menggunakan metode klasifikasi Naive Bayes menghasilkan nilai sebesar 86,56% dan nilai akurasi yang diperoleh dengan menggunakan metode K-NN menghasilkan nilai sebesar 74,78% [16].

Berdasarkan penelitian terkait diatas dapat dijadikan referensi dalam melakukan penelitian ini. Penelitian ini menggunakan algoritma Naive Bayes dan K-NN untuk menghasilkan nilai akurasi dalam klasifikasi sentimen analisis pengguna twitter terkait vaksin booster dengan kata kunci "vaksin booster" dalam sentimen positif, dan negatif.

II. METODE PENELITIAN

Penelitian ini menggunakan data yang diperoleh melalui twitter. Data yang di diambil menggunakan tools Rapidminer. Adapun tahapan penelitian menggunakan tahapan CRIPS-DM yang merupakan metode dengan memanfaatkan model proses yang banyak digunakan para ahli dalam pengembangan data untuk memecahkan suatu masalah. Ada 6 tahapan dalam CRIPS-DM yaitu tahap pemahaman bisnis, tahap pemahaman data, tahap persiapan data, tahap pemodelan, tahap evaluasi dan tahap penyebaran.



Gambar 1. CRIPS-DM

1. Pemahaman Bisnis (Business Understanding)

Pemahaman bisnis atau business understanding adalah proses awal dalam CRIPS-DM yang merupakan pemahaman tentang tujuan kegiatan penelitian. Dalam penelitian ini membutuhkan pengetahuan untuk mendapatkan data yang mendukung proses analisis sentimen pengguna twitter terkait vaksin booster.

2. Pemahaman Data (Data Understanding)

Pemahaman data atau data understanding merupakan proses pengumpulan data twitter dengan melakukan crawling data pada twitter dengan menggunakan kata kunci vaksin booster serta mengidentifikasi dan memahami data yang dimiliki.

3. Persiapan Data (Data Preparation)

Pada proses persiapan data atau data preparation ini telah didapatkan data dari crawling twitter sebelumnya



kemudian data akan dilakukan data processing. Berikut adalah langkah-langkah dalam data processing :

- Transform case merupakan tahapan mengubah data menjadi lower case atau huruf kecil
- Tokenizing merupakan tahapan pemecahan karakter dalam suatu teks kedalam satuan kata dan menghilangkan kata, simbol, karakter unik pada dataset.
- Filter Tokens (by Length) merupakan tahapan menghilangkan sejumlah kata dengan panjang karakter tertentu.
- Stopwords merupakan tahapan menghilangkan kata yang tidak memiliki makna.
- Stemming merupakan tahapan pencarian kata dasar dari setiap kata.

4. Pemodelan (Modelling)

Tahapan pemodelan dilakukan dengan menggunakan teknik pemodelan yang digunakan pada dataset yang telah disiapkan. Teknik pemodelan dalam penelitian ini menggunakan algoritma klasifikasi Naïve Bayes dan K-Nearst Neighbor. Tools yang digunakan dalam tahapan pemodelan ini adalah Rapidminer.

5. Evaluasi (Evaluation)

Tahapan evaluasi ini dengan menguji model yang telah diterapkan dan melakukan evaluasi metode klasifikasi dengan menggunakan confusion matrix terhadap algoritma Naïve Bayes dan K-Nearst Neighbor.

6. Penyebaran (Deployment)

Tahapan ini akan dilakukan pemberian kesimpulan yang telah didapatkan dalam melakukan penelitian ini.

preprocessing. Proses preprocessing data tweet dibutuhkan karena beberapa tweet tidak menggunakan kata yang baku, banyak menggunakan simbol atau karakter unik. Tahapan proses preprocessing sebagai berikut:

- Replace
Proses menghilangkan data tweet yang mengandung seperti link, hastag (#), titik(.), koma(,) dan karakter
- Transform Cases
Proses transform cases pada Rapidminer digunakan untuk mengubah kalimat pada tweet menjadi huruf kecil atau lower case
- Tokenize
Tahapan pemecahan karakter dalam suatu teks kedalam satuan kata atau bisa juga menghilangkan kata yang berisi angka, simbol dan karakter unik lainnya.
- Stopword
Tahapan menghilangkan kata yang tidak memiliki makna atau tidak diperlukan dalam menghilangkan kata- kata dalam Bahasa Indonesia yang tidak relevan seperti kata “dan”.
- Filter (by Length)
Tahapan ini berfungsi untuk menghilangkan sejumlah kata dengan panjang karakter tertentu.

Pemodelan (Modelling)

Tahapan pemodelan ini dilakukan dengan memanfaatkan teknik algoritma klasifikasi Naïve Bayes dan teknik algoritma klasifikasi K-Nearst Neighbor (K-NN) yang digunakan untuk mengklasifikasi sentimen.

III. HASIL DAN PEMBAHASAN

Pemahaman Bisnis (Bussiness Understanding)

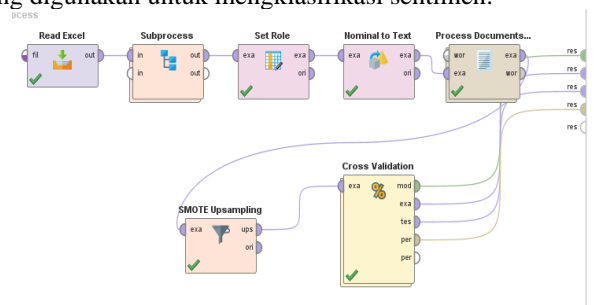
Tahapan pemahaman bisnis atau business understanding merupakan pemahaman tujuan berdasarkan kegiatan penelitian sentimen analisis vaksin booster yang akan dilakukan. Tujuan dilakukannya penelitian mengenai analisis sentimen terkait vaksin booster dengan Analisis sentimen ini dilakukan untuk menemukan metode klasifikasi dengan menggunakan metode Naïve Bayes dan K-Nearst Neighbor (K-NN)

Pemahaman Data (Data Understanding)

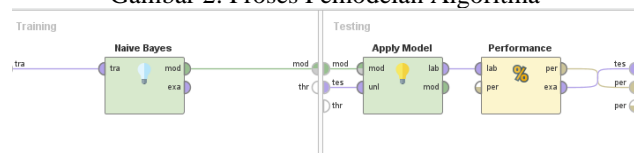
Tahapan pemahaman data atau data understanding merupakan tahap mengumpulkan data di media sosial twitter menggunakan dengan menggunakan kata kunci “vaksin booster” yang digunakan untuk menjumlahkan dataset di Twitter selama periode 4 Juni-11 Juni 2022 . Data yang didapatkan dari hasil crawling data selama periode 4 Juni-11 Juni 2022 sebanyak 1000 data.

Persiapan Data (Data Preparation)

Pada tahapan ini data yang telah dikumpulan dari hasil crawling selanjutnya data tersebut akan dilakukan data

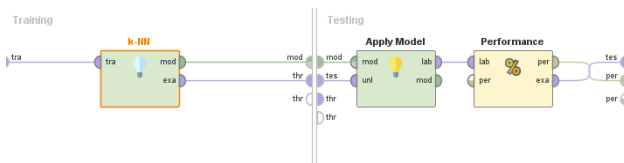


Gambar 2. Proses Pemodelan Algoritma



Gambar 3. Pengujian Model Naïve Bayes

Dalam proses Cross Validation data menggunakan metode klasifikasi algoritma Naïve Bayes yang meliputi proses training dan testing menggunakan apply model dan performance. Setelah itu dilakukan proses pemodelan dengan menggunakan metode klasifikasi algoritma Naïve Bayes untuk mendapatkan nilai accuracy, Precision, dan recall berdasarkan performance vector.



Gambar 4. Pengujian Model K-Nearst Neighbor

Dalam proses Cross Validation menggunakan metode klasifikasi algoritma K-Nearst Neighbor yang meliputi proses training dan testing dengan apply model dan performance. Setelah itu dilakukan proses pemodelan dengan menggunakan metode klasifikasi algoritma K-Nearst Neighbor untuk mendapatkan nilai accuracy, precision, dan recall berdasarkan performance vector. Evaluasi (Evaluation)

Evaluasi merupakan tahap untuk melihat hasil dari model yang telah dibuat menggunakan Confusion Matrix serta menggunakan Cross Validation. Cross Validation menggunakan 10 fold Cross Validation adalah salah satu k-fold Cross Validation yang banyak direkomendasikan dalam pengujian karena memberikan estimasi akurasi yang baik. Berikut hasil accuracy dari algoritma Naïve Bayes dan K-Nearst Neighbor :

accuracy: 90.19% +/- 2.83% (micro average: 90.19%)

	true Positif	true Negatif	class precision
pred. Positif	539	27	95.23%
pred. Negatif	98	610	86.16%
class recall	84.62%	95.76%	

Gambar 5. Hasil Accuray Algoritma Naïve Bayes

Hasil pada metode Naïve Bayes menyatakan bahwa tingkat akurasi pada metode ini sebesar 90.19%. Dapat dilihat bahwa class precesion untuk pred.positif adalah 95.23% dan pred.negatif adalah 86.16%.

accuracy: 82.89% +/- 1.72% (micro average: 82.89%)

	true Positif	true Negatif	class precision
pred. Positif	625	206	75.21%
pred. Negatif	12	431	97.29%
class recall	98.12%	67.66%	

Gambar 6. Hasil Accuracy K-Nearst Neighbor

Hasil pada metode Naïve Bayes menyatakan bahwa tingkat akurasi pada metode ini sebesar 82.89% . Dapat dilihat bahwa class precesion untuk pred.positif adalah 75.21% dan pred.negatif adalah 97.29%.

Tabel 1. Perbandingan Hasil Accuracy

Algoritma	Hasil Accuracy
Naïve Bayes	90.19%
K-Nearst Neighbor	82.89%

IV. PENUTUP

Berdasarkan hasil yang didapat setelah penelitian ini, maka dapat diambil kesimpulan yaitu berdasarkan hasil pengujian dan klasifikasi analisis sentiment pengguna twitter terhadap vaksin booster sebanyak 1000 data didapatkan data kelas label positif sebanyak 637 tweet dan label negatif sebanyak 363 tweet menggunakan dua metode yaitu klasifikasi algoritma Naïve Bayes dan klasifikasi algoritma K-Nearst Neighbor (K-NN) memiliki hasil akurasi yang berbeda. Hasil akurasi yang didapatkan dengan metode klasifikasi Naïve Bayes adalah sebesar 90.19% lebih baik dibandingkan dengan metode klasifikasi algoritma K-Nearst Neighbor (K-NN) yaitu sebesar 82.89%. Nilai precision menggunakan metode Naïve Bayes dengan pred.positif lebih tinggi sebesar 95.23% dibandingkan pred.negatif sebesar 86.16%.

V. REFERENSI

[1] “Peta Sebaran Covid19.go.id.” <https://covid19.go.id/peta-sebaran> (accessed Apr. 18, 2022).

[2] “Vaksin Dashboard.” <https://vaksin.kemkes.go.id/#/vaccines> (accessed Apr. 19, 2022).

[3] D. Hernikawati, “Kecenderungan Tanggapan Masyarakat Terhadap Vaksin Sinovac Berdasarkan Lexicon Based Sentiment Analysis The Trend of Public Response to Sinovac Vaccine Based on Lexicon Based Sentiment Analysis,” vol. 23, no. 1, pp. 21–31, 2021.

[4] A. Suwarno and A. Andriani, “Analisis Sentimen Pada Media Sosial Twitter Mengenai Tanggapan Vaksinasi COVID-19 Menggunakan Metode Naive Bayes,” *J. Tek. Ind.*, 2021, Accessed: Apr. 11, 2022. [Online]. Available: <https://jurnal.pelitabangsa.ac.id/index.php/JUTIN/article/view/906>.

[5] F. Fathonah and A. Herliana, “Penerapan Text Mining Analisis Sentimen Mengenai Vaksin Covid - 19 Menggunakan Metode Naïve Bayes,” vol. 7, no. November, pp. 155–164, 2021, doi: 10.34128/jsi.v7i2.331.

[6] M. I. Aditama, R. I. Pratama, K. Hafizzana, U. Wiwaha, and N. A. Rakhmawati, “Analisis Klasifikasi Sentimen Pengguna Media Sosial Twitter Terhadap Pengadaan Vaksin COVID-19,” vol. 04, pp. 90–92, 2020.

[7] D. Darwis, N. Siskawati, and Z. Abidin, “Penerapan Algoritma Naive Bayes untuk Analisis Sentimen Review Data Twitter BMKG Nasional,” vol. 15, no. 1, pp. 131–145.

[8] A. Baita, Y. Pristyanto, N. Cahyono, P. Covid-, K. N. N. Akurasi, and K. Kunci, “ANALISIS SENTIMEN MENGENAI VAKSIN SINOVAK MENGGUNAKAN ALGORITMA SUPPORT VECTOR MACHINE (SVM) DAN K-NEAREST NEIGHBOR (KNN) Abstraksi Keywords ;,” vol. 4, no. 2, pp. 42–46, 2021.

[9] R. Fahlapi *et al.*, “ANALISA SENTIMEN VAKSINASI COVID-19 DENGAN METODE SUPPORT VECTOR MACHINE DAN NAÏVE BAYES BERBASIS,” vol. 6, no. 1, pp. 57–64, 2022.

[10] K. Bhuana and L. Muflikhah, “Analisis Sentimen Masyarakat Indonesia tentang Vaksin Covid-19 di Twitter dengan menggunakan Metode K- Nearest



- Neighbors dan Seleksi Fitur Chi Square,” vol. 6, no. 3, pp. 1395–1401, 2022.
- [11] S. Lestari and S. Saepudin, “Analisis Sentimen Vaksin Sinovac Pada Twitter Menggunakan Algoritma Naive Bayes,” *SISMATIK (Seminar Nas. Sist. Inf. dan Manaj. Inform.*, pp. 163–170, 2021.
- [12] S. Ernawati and R. Wati, “Penerapan Algoritma K-Nearest Neighbors Pada Analisis Sentimen Review Agen Travel,” *J. Khatulistiwa Inform.*, vol. 6, no. 1, pp. 64–69, 2018.
- [13] T. T. Widowati *et al.*, “ANALISIS SENTIMEN TWITTER TERHADAP TOKOH PUBLIK DENGAN ALGORITMA NAIVE BAYES DAN SUPPORT VECTOR,” vol. 11, no. 2, 2020.
- [14] M. S. Alrajak, I. Ernawati, and I. Nurlaili, “ANALISIS SENTIMEN TERHADAP PELAYANAN PT PLN DI JAKARTA PADA TWITTER DENGAN ALGORITMA K- NEAREST,” pp. 110–122, 2020.
- [15] J. S. Wibowo, “ANALISIS SENTIMEN KLASIFIKASI TWEET VAKSIN COVID 19 DENGAN NAÏVE BAYES,” vol. 6, no. 2, 2021.
- [16] J. Jtik, J. Teknologi, S. H. Ramadhani, and M. I. Wahyudin, “Analisis Sentimen Terhadap Vaksinasi Astra Zeneca pada Twitter Menggunakan Metode Naïve Bayes dan K-NN,” vol. 6, no. 4, 2022.