

Sentimen Analisis Pandangan Masyarakat Terhadap Vaksinasi Covid 19 Menggunakan K-Nearest Neighbors

Dyah Apriliani^{1*}, Ardi Susanto², Muhammad Fikri Hidayattullah³, Ginanjar Wiro Sasmito⁴

^{1,2,3,4}Program Studi Teknik Informatika, Politeknik Harapan Bersama,

^{1,2,3,4}Jalan Mataram No. 09, Kota Tegal, 50272, Indonesia

email: ¹dyahapril@poltektegal.ac.id, ²ardisusanto@poltektegal.ac.id, ³fikri@poltektegal.ac.id, ⁴ginanjar@poltektegal.ac.id

Abstract – The Covid 19 pandemic that occurred was very disturbing to the community. Many people have been exposed to or have lost their families because of this virus. To prevent further spread of the covid 19 virus, the government is holding a vaccination program. The vaccination program that was carried out reaped pros and cons from the community. Based on these problems, this research will carry out the process of classifying people's views on the Covid 19 vaccination. The research data was taken from Twitter as many as 2241. The data will be classified into two classes, namely positive and negative. The classification process will be carried out using the K-Nearest Neighbors (KNN) method. The first stage is data collection from Twitter, data labeling, data preprocessing with Python, TF-IDF weighting, model building, model testing, and model evaluation. This study obtains the best accuracy of 79.25% using K-Fold 10 and KNN 5 parameters.

Abstrak – Pandemi covid 19 yang terjadi sangat meresahkan masyarakat. Banyak masyarakat yang terpapar maupun kehilangan keluarga mereka karena virus ini. Untuk mencegah semakin menyebarnya virus covid 19, pemerintah menyelenggarakan program vaksinasi. Program vaksinasi yang dilakukan menuai pro dan kontra dari masyarakat. Berdasarkan permasalahan tersebut, maka dalam penelitian ini akan melakukan proses klasifikasi pandangan masyarakat terhadap vaksinasi Covid 19. Data penelitian yang digunakan diambil dari twitter sebanyak 2241 data. Data akan diklasifikasikan menjadi 2 kelas yaitu positif dan negatif. Proses klasifikasi akan dilakukan dengan menggunakan metode *K-Nearest Neighbors* (KNN). Tahapan pertama yang dilakukan adalah pengambilan data dari twitter, pelabelan data, preprocessing data di python, pembobotan TF-IDF, pembuatan model, pengujian model dan evaluasi model. Dari penelitian ini didapatkan akurasi terbaik sebesar 79,25% dengan menggunakan parameter K-Fold 10 dan KNN 5.

Kata Kunci – Analisis sentimen, vaksinasi, covid 19, KNN

I. PENDAHULUAN

Awal tahun 2020 menjadi awal munculnya wabah penyakit yang penyebarannya sangat pesat, yang melanda lebih dari 190 negara. Wabah tersebut dikenal dengan covid-19 yang penyebabnya adalah virus SARS-Cov (Servere Adute Respiratory Syndrome Coronavirus-2). Penyebaran wabah penyakit tersebut semakin marak, sehingga sejak tanggal 12 Maret 2020, WHO menetapkan wabah ini menjadi pandemi.

Perkembangan wabah yang semakin meluas dan bahaya yang ditimbulkan dari penyakit tersebut, maka dilakukan berbagai upaya untuk pencegahan penyebarannya. Salah satu

upaya yang dilakukan adalah dengan melakukan pengembangan vaksin untuk penyakit ini[1]. Tujuan adanya vaksin yaitu untuk melindungi dan mencegah maraknya penularan virus ini. Vaksinasi yang dilakukan diharapkan mampu memberikan perlindungan dan mengurangi penyebaran virus di lingkungan sekitar[2]. Pemerintah menganjurkan masyarakat untuk mengikuti vaksinasi. Di Indonesia, pemerintah dalam rangka mencegah penyebaran wabah penyakit ini juga mengeluarkan aturan melalui Peraturan Presiden Nomor 99 Tahun 2020 Tentang Pengadaan vaksin dan pelaksanaan vaksinasi dalam rangka penanggulangan pandemi Corona Virus Disease 2019 (Covid-19). Peraturan tersebut dibuat untuk mengatur kewenangan pemerintah, kementerian/lembaga dan para pejabatnya dalam rencana kegiatan vaksinasi [3].

Pelaksanaan vaksinasi yang dicanangkan oleh pemerintah diharapkan mampu menanggulangi penyebaran dari virus ini. Dengan adanya kebijakan tersebut, menimbulkan banyak asumsi yang menyebar di masyarakat. Program vaksinasi tersebut juga menimbulkan pro dan kontra dari masyarakat. Ungkapan maupun asumsi masyarakat ini juga banyak disampaikan melalui sosial media yang menjadi perbincangan publik [4]. Media komunikasi publik yang sering digunakan adalah twitter. Kumpulan opini masyarakat yang diungkapkan dalam sosial media ini mengandung banyak opini sentimen.

Analisis teks dapat dilakukan untuk menghasilkan informasi spesifik yang dibutuhkan[5]. Dari opini masyarakat yang muncul di sosial media tersebut dapat dilakukan analisis untuk mengetahui pandangan masyarakat mengenai vaksinasi covid-19. Klasifikasi yang dilakukan dengan pengambilan data di twitter, dapat diambil secara otomatis melalui postingan dan tweet dari pengguna. Klasifikasi dilakukan menjadi beberapa kelas [6]. Berdasarkan permasalahan tersebut, maka akan dilakukan klasifikasi pandangan masyarakat.

Pertama data akan diklasifikasikan menjadi 2 label yaitu positif dan negatif. Dari hasil klasifikasi tersebut, dapat digunakan oleh masyarakat untuk menentukan keputusan apakah komentar tersebut masuk ke kategori positif atau negatif [7]. Klasifikasi pandangan masyarakat ini akan diproses dengan menggunakan metode *K-Nearest Neighbors* (KNN). Metode KNN mampu menghasilkan tingkat akurasi yang baik untuk proses klasifikasi. Dalam penelitian ini dilakukan beberapa tahapan, diantaranya: pengambilan data dari twitter, pelabelan data, preprocessing data, pembobotan dengan TF-IDF, pembuatan model, pengujian dan evaluasi dari model yang didapatkan.

*) penulis korespondensi: Dyah Apriliani

Email: dyahapril@poltektegal.ac.id

II. PENELITIAN YANG TERKAIT

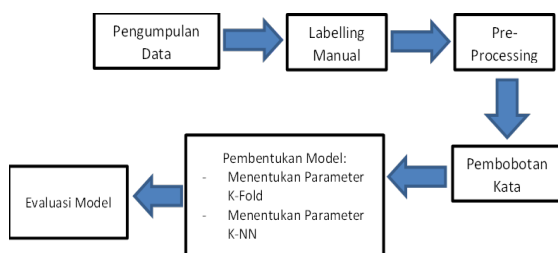
Metode Improve K-Nearest Neighbor adalah metode klasifikasi. Dalam penelitian yang dilakukan oleh Herdiawan, metode *K-Nearest Neighbors* (KNN) ini digunakan untuk membagi opini masyarakat terhadap layanan penyedia internet yaitu Telkom Indihome. Opini masyarakat diklasifikasikan dalam dua kelas, yaitu positif dan negatif. Dataset diperoleh dari data twitter yang berasal dari data komentar layanan baik itu komentar yang bagus maupun yang sebaliknya. Tahapan yang dilakukan dalam penelitian ini yaitu: preprocessing data, pembobotan kata, dan dilanjutkan dengan pengkategorian. Metode KNN memberikan hasil klasifikasi yang baik. Dalam penelitian ini didapatkan tingkat akurasi sebesar 80%. Hasil akurasi yang dipengaruhi oleh beberapa faktor yaitu data set yang digunakan, preprocessing yang pemilihan parameter yang digunakan [8].

Metode *K-Nearest Neighbors* (KNN) juga digunakan dalam analisis sentimen untuk review Objek Wisata. Tujuan dilakukan review ini yaitu agar memudahkan pengunjung untuk mengetahui informasi mengenai tempat wisata yang akan di datangi. Penggunaan metode KNN disini dikarenakan metode ini memiliki prinsip yang sederhana, yaitu data diolah dengan menghitung jarak terpendek dari data *training* dan data *testing*. Dari penelitian ini diperoleh tingkat akurasi sebesar 77,01%, dengan nilai $k=7$ [9].

Penelitian lain yang membahas tentang sentimen analisis vaksin covid 19, yaitu dengan memanfaatkan metode Naïve Bayes [10],[11], dan juga dengan menggunakan metode Naïve Bayer, *Random Forest* dan *Support Vektor Machine SVM* [12]. Dari penelitian tersebut data diambil dari media sosial twitter. Penelitian ini menghasilkan tingkat akurasi yang berbeda-beda. Dikarenakan data yang diambil dari twitter memiliki rentan waktu yang berbeda. Penelitian menggunakan metode *K-Nearest Neighbors*.

III. METODE PENELITIAN

Penelitian ini dilakukan untuk mengetahui pandangan masyarakat tentang vaksinasi Covid-19 yang diklasifikasi memakai metode *K-Nearest Neighbors*. Tahapan penelitian dilakukan dengan beberapa tahapan, diantaranya: penyumpulan data, labeling manual, Pre-Processing, Pembobotan kata, pembentukan model dan evaluasi model. Adapun tahapan dari penelitian yang dilakukan bisa dilihat pada Gambar 3.1 di bawah ini.



Gambar 3.1. Prosedur Penelitian

Prosedur penelitian yang dilakukan dapat dijelaskan secara detail sebagai berikut:

1. Pengumpulan Data
Data yang digunakan dalam penelitian ini diperoleh dari crawling data di twitter dengan menggunakan kata kunci “Vaksin Booster”. Dari tahapan tersebut diperoleh data sebanyak 2.241. Dari data yang telah dihasilkan dilanjutkan pemberian label data dikerjakan secara manual.

2. Labeling Manual
Setelah data terkumpul tahap selanjutnya yaitu dengan melakukan pelabelan manual. Pelabelan dilakukan dengan mengklasifikasikan komentar kedalam dua kelas, yaitu positif dan negatif. Cara kerja dari pelabelan ini yaitu dengan memahami makna dari kalimat, bukan dilakukan per kata. Pada Tabel 3.1. dibawah ini adalah panduan proses pelabelan.

Tabel 31. Panduan Proses Pelabelan

No	Kelas	Kondisi
1	Positif	Data berupa ungkapan dukungan, saran, cinta, doa, pertanyaan tentang vaksin, jadwal vaksin dan kuota Vaksin
2	Negatif	Data berupa ungkapan ketidakpercayaan, fitnah, hoax, juga hal-hal buruk dari vaksinasi itu sendiri baik yang berupa sindiran maupun pernyataan yang tidak setuju dengan vaksin covid-19

3. Pre-Processing
Tahapan selanjutnya yang dilakukan setelah melakukan pelabelan adalah *pre-processing* data. Tujuan dari tahapan ini adalah melakukan pengolahan dataset agar data dapat terbaca oleh algoritma yang akan digunakan, dan juga melakukan ekstraksi informasi yang diinginkan. Tahapan yang dilakukan dalam pre-processing data yaitu:
 - a. Menghapus semua karakter khusus seperti emotikon.
 - b. Menghapus spasi ganda maupun tanda baca lainnya.
 - c. Mengubah semua data ke dalam huruf kecil
 - d. Menghapus stopwords, yaitu kata yang sering muncul, diantaranya: adalah, adanya, adapun dan yang lainnya.

4. Pembobotan Kata
Pembobotan kata yang dilakukan yaitu dengan menggunakan metode TF-IDF. Metode ini merupakan salah satu cara yang digunakan untuk memberikan pembobotan pada kata dengan suatu dokumen. Bobot dalam metode ini ditentukan berdasarkan frekuensi kemunculan kata dalam dokumen yang digunakan.

5. Pembentukan Model
Pembentukan model dilakukan dengan menggunakan metode KNN. Data yang telah dilakukan pelabelan, pre-processing dan pembobotan selanjutnya dibagi menjadi data latih dan data uji. Pembagian yang dilakukan yaitu 70% digunakan untuk data latih sedangkan 30% digunakan untuk data uji. Setelah model terbentuk, selanjutnya dilakukan pengujian

dengan menentukan parameter cross validation dan parameter KNN yang akan digunakan. Pada Tabel 3.2 di bawah ini dapat dilihat parameter pengujian yang digunakan dalam penelitian ini.

Tabel 3.2 Parameter Pengujian

No	K-Fold	KNN
1	10	2
2	10	3
3	10	4
4	10	5

6. Evaluasi Model

Tahapan terakhir yang dilakukan yaitu dengan melakukan evaluasi model yang telah dihasilkan. Evaluasi model dilakukan dengan menghitung akurasi. Perhitungan akurasi yaitu perhitungan yang didapatkan dari perhitungan jumlah total prediksi yang benar. Perhitungan akurasi dihitung dengan menggunakan *Confusion Matrix*. Dalam persamaan 3.1 dibawah ini adalah persamaan *confusion matrix* [13][14][15].

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \quad (3.1)$$

Dimana:

TP = True Positive (hasil prediksi positif dan nilai sebenarnya bernilai positif)

TN = True Negative (hasil prediksi negatif dan nilai sebenarnya bernilai negatif)

FP = False Positive (hasil prediksi positif dan nilai sebenarnya bernilai negatif)

FN = False Negative (hasil prediksi negatif dan nilai sebenarnya bernilai positif)

Selain menghitung akurasi dalam evaluasi juga dilakukan perhitungan *precision*. *Precision* digunakan untuk mengetahui ketepatan informasi yang diinginkan pengguna dari penerapan parameter yang diujikan. Adapun rumus yang digunakan untuk menghitung *Precision* dapat dilihat pada persamaan 3.2.

$$Precision = \frac{TP}{TP + FP} \quad (3.2)$$

Perhitungan juga dilakukan untuk menghitung *recall*. *Recall* digunakan untuk mengetahui jumlah kesesuaian yang terklasifikasi dengan hasil yang didapatkan oleh pengguna. Rumus *recall* yang digunakan dapat dilihat dalam persamaan 3.3 di bawah ini.

$$Recall = \frac{TP}{TP + FN} \quad (3.3)$$

Precision dan *Recall* merupakan salah satu pendukung penting dalam perhitungan *Confusion Matrix*. Keduanya dapat digunakan untuk melakukan perbaikan algoritma yang digunakan.

IV. HASIL DAN PEMBAHASAN

Penelitian yang dilakukan menghasilkan suatu model terbaik dari metode *K-Nearest Neighbors* yang digunakan

untuk melakukan sentimen analisis terhadap pandangan masyarakat mengenai vaksin Booster Covid 19 di Indonesia. Pengambilan data yang dilakukan di twitter mendapatkan data sebanyak 2241 data. Dari data tersebut dibagi menjadi dua kelas yaitu 1370 data kelas positif dan 871 data kelas negatif. Penelitian dilakukan dengan melakukan training dan testing dengan menggunakan beberapa parameter. Pada Tabel 1 dibawah ini adalah pengujian model yang dilakukan dengan menggunakan parameter K-Fold 10 dan KNN 2. Dengan menggunakan parameter ini, hasil akurasi yang didapatkan adalah sebesar 77,96%.

Tabel 4.1. Pengujian dengan K-Fold 10 dan KNN 2

	True Positif	True Negative	Class precision
Pred. Positive	1128	252	81,74%
Pred. Negative	242	619	71,89%
Class recall	82,34%	71,07%	

Pengujian kedua dilakukan dengan menggunakan parameter lain, yaitu dengan menggunakan K-Fold 10 dan KNN 3. Berdasarkan parameter ini dihasilkan nilai akurasi sebesar 78,50%. Hasil pengujian dapat dilihat pada Tabel 4.2

Tabel 4.2. Pengujian dengan K-Fold 10 dan KNN 3

	True Positif	True Negative	Class precision
Pred. Positive	1140	254	81,78%
Pred. Negative	230	617	72,85%
Class recall	83,21%	70,84%	

Pengujian ketiga dilakukan dengan menggunakan parameter yang berbeda lagi, yaitu dengan menggunakan K-Fold 10 dengan parameter KNN 4. Berdasarkan hasil pengujian tersebut didapatkan nilai akurasi sebesar 78,63%. Hasil akurasi ini lebih tinggi dibandingkan dengan pengujian yang dilakukan sebelumnya. Pada Tabel 4.3 adalah hasil pengujian menggunakan parameter tersebut.

Tabel 4.3. Pengujian dengan K-Fold 10 dan KNN 4

	True Positif	True Negative	Class precision
Pred. Positive	1145	254	81,84%
Pred. Negative	225	617	73,28%
Class recall	83,58%	70,84%	

Pengujian terakhir dengan menggunakan parameter ini dilakukan dengan melakukan perubahan parameter yaitu dengan menggunakan parameter K-Fold 10 dan KNN 5. Tingkat akurasi yang dihasilkan dengan menggunakan parameter ini yaitu sebesar 79,25%. Dari hasil pengujian yang dilakukan maka pemilihan parameter ini yang menghasilkan nilai kaurasi yang tertinggi dibandingkan dengan beberapa parameter yang telah digunakan. Adapun hasil pengujian secara details dapat dilihat pada Tabel 4.4 dibawah ini.

Tabel 4.4. Pengujian dengan K-Fold 10 dan KNN 4

	True Positif	True Negative	Class precision
Pred. Positive	1143	238	82,77%
Pred. Negative	227	633	73,60%
Class recall	83,43%	72,68%	

Selain menggunakan metode KNN, pengujian dilakukan menggunakan metode lain yaitu menggunakan Naive Bayes dan *Support Vector Machine* (SVM). Pengujian dilakukan untuk mengetahui perbandingan nilai akurasi yang dihasilkan. Pengujian pertama dilakukan dengan menggunakan parameter K-Fold 10 dan digabungkan dengan menggunakan metode Naive Bayes. Hasil akurasi yang didapatkan dengan menggunakan parameter tersebut adalah

sebesar 65,64%. Pada Tabel 4.5 dibawah ini adalah hasil pengujian untuk parameter tersebut.

Tabel 4.5. Pengujian dengan K-Fold 10 dan Naïve Bayes

	True Positif	True Negative	Class precision
Pred. Positive	759	146	83,87%
Pred. Negative	634	731	53,55%
Class recall	54,49%	83,35%	

Pengujian juga dilakukan dengan menerapkan teknik lain, yaitu dengan menggunakan metode SVM dan K-Fold 10. Berdasarkan hasil pengujian dengan parameter ini maka didapatkan tingka akurasi sebesar 75,20%. Pada Tabel 4.6 adalah hasil dari pengujian dengan menggunakan parameter K-Fold 10 dan metode SVM.

Tabel 4.6. Pengujian dengan K-Fold 10 dan Metode SVM

	True Positif	True Negative	Class precision
Pred. Positive	1373	543	71,66%
Pred. Negative	20	334	94,35%
Class recall	98,56%	38,08%	

Penelitian telah dilakukan dan diujicoba dengan menggunakan beberapa metode, diantaranya *K-Nearest Neighbors (KNN)*, *Naïve Bayes* maupun metode *Support Vektor machine (SVM)*. Berdasarkan hasil yang telah dilakukan maka dapat dilihat bahwa tingkat akurasi tertinggi diperoleh dengan menggunakan metode KNN dengan parameter yang digunakan k = 5 dan K-Fold 10, yaitu dengan nilai akurasi sebesar 79,25%.

V. KESIMPULAN

Dalam penelitian ini, dilakukan sentimen analisis terhadap pandangan masyarakat terhadap vaksin Booster Covid 19 dengan menggunakan metode *K-Nearest Neighbors*. Nilai akurasi tertinggi yang didapatkan dari model yang dihasilkan adalah sebesar 79,25% dengan menggunakan parameter K-Fold 10 dan KNN 5. Nilai akurasi yang dihasilkan dari model ini masih sangat kecil sehingga perlu adanya penambahan parameter yang lebih kompleks. Preprocessing juga mempengaruhi tingkat akurasi sehingga dalam pengembangan selanjutnya dapat mengganti atau mencoba metode preprocessing yang lain, misalnya Word2Vektor. Selain itu juga bisa dilakuakn penambahan data baik itu data yang positif maupun data yang negatif.

UCAPAN TERIMA KASIH

Penulis berterimakasih untuk semua pihak yang telah memberikan bantuan ataupun memberi *support* terkait dengan penelitian ini, terutama untuk Politeknik Harapan Bersama.

DAFTAR PUSTAKA

[1] C. Liu *et al.*, “Research and Development on Therapeutic Agents and Vaccines for COVID-19 and Related Human Coronavirus Diseases,” *ACS Cent. Sci.*, vol. 6, no. 3, pp. 315–331, Mar. 2020, doi: 10.1021/acscentsci.0c00272.

[2] I. P. Sari and S. Sriwidodo, “Perkembangan Teknologi Terkini dalam Mempercepat Produksi Vaksin COVID-19,” *Maj. Farmasetika*, vol. 5, no. 5, p. 204, Aug. 2020, doi:

10.24198/mfarmasetika.v5i5.28082.

[3] “PERATURAN PRESIDEN. REPUBLIK INDONESIA,” pp. 1–13, 2020.

[4] N. Anggraini and H. Suroyo, “Comparison of Sentiment Analysis against Digital Payment ‘T-cash and Go-pay’ in Social Media Using Orange Data Mining,” *J. Inf. Syst. Informatics*, vol. 1, no. 2, pp. 152–163, Sep. 2019, doi: 10.33557/journalisi.v1i2.21.

[5] E. S. Negara, R. Andryani, and P. H. Saksono, “Analisis Data Twitter: Ekstraksi dan Analisis Data Geospasial,” *J. INKOM*, vol. 10, no. 1, p. 27, Nov. 2016, doi: 10.14203/j.inkom.433.

[6] A. Tripathy, A. Agrawal, and S. . Rath, “Classification of sentiment reviews using n-gram machine learning approach,” *Expert Syst. Appl.*, vol. 57, pp. 117–126, 2016.

[7] M. Adrian, J. Asian, B. Nazief, and H. E. Williams, “Stemming Indonesian: A Confix-Stripping Approach,” *ACM Transactions on Asian Language Information Processing*, vol. 6, no. 4, pp. 1–33, 2007.

[8] Herdiawan, “Analisis Sentimen Terhadap Telkom Indihome Berdasarkan Opini Publik Menggunakan Metode Improved K-Nearest Neighbor,” 2013, [Online]. Available: <http://elib.unikom.ac.id/gdl.php?mod=browse&op=read&id=jbptunikompp-gdl-herdiawann-33861>.

[9] R. Sari, “ANALISIS SENTIMEN PADA REVIEW OBJEK WISATA DUNIA FANTASI MENGGUNAKAN ALGORITMA K-NEAREST NEIGHBOR (K-NN),” *EVOLUSI J. Sains dan Manaj.*, vol. 8, no. 1, Mar. 2020, doi: 10.31294/evolusi.v8i1.7371.

[10] R. Al Habsi, R. A. D. Anggoro, and M. A. Valio, “ANALISIS SENTIMEN TERHADAP VAKSIN COVID-19 DI JEJARING SOSIAL TWITTER MENGGUNAKAN ALGORITMA NAÏVE BAYES,” 2021.

[11] Z. Firmansyah and N. F. Puspitasari, “ANALISIS SENTIMEN MASYARAKAT TERHADAP VAKSINASI COVID-19 BERDASARKAN OPINI PADA TWITTER MENGGUNAKAN ALGORITMA NAIVE BAYES,” *JTI*, vol. 14, pp. 171–178, 2021, doi: 10.15408/jti.v14i2.24024.

[12] A. F. Nuraini, R. D. Pertiwi, M. Z. Subarkah, and K. Ferawati, “Analisis Sentimen Pelaksanaan Vaksinasi Covid-19 secara Massal pada Media Sosial Twitter,” pp. 1385–1394.

[13] A. Anggraeni and I. Cholissodin, “Sentimen Analisis Layanan Produk Indihome menggunakan Information Gain dan Metode K-Nearest Neighbor,” vol. 5, no. 8, pp. 3616–3624, 2021.

[14] I. W. Saputro and B. W. Sari, “Uji Performa Algoritma Naïve Bayes untuk Prediksi Masa Studi Mahasiswa,” *Citec J.*, vol. 6, pp. 1–11, 2019.

[15] H. Zidny Ilmadina, D. Apriliani, and D. S. Wibowo, “Deteksi Pengendara Mengantuk dengan Kombinasi Haar Cascade Classifier dan Support Vector Machine,” vol. 7, no. 1, pp. 1–7, 2022.